Emotive aspects of face perception and the human brain

Joel S. Winston

2005

A thesis submitted to University College London
for the degree of Doctor of Philosophy

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

I would like to thank everyone at the Wellcome Department of Imaging Neuroscience for their help and support. In particular, Paul Bentley, Jay Gottfried and James Kilner were a constant source of distraction and inspiration, in approximately the right balance. Miriam Fine-Goulden kindly assisted with the collection of behavioural data for the subsidiary experiment reported in Chapter 3. A number of people helped with the provision or production of stimuli that were used in the experiments, including Jan-Eric Litton, Oliver Josephs and Masakazu Fujimiya. I was fortunate enough to collaborate with a number of talented scientists in producing the work for this thesis, and I thank Rik Henson, John O'Doherty, Patrik Vuilleumier, Bryan Strange and David Perrett for their time and effort. Ray Dolan has been the ideal supervisor: guiding but never dogmatic. A number of people have helped by proofreading sections of this manuscript; in particular, I thank Robert Winston and Karin Shmueli. Finally, I thank my family and friends for their encouragement and interest; in particular, I thank Joanne for absolutely everything.

Abstract

The neural mechanisms by which faces are processed are the subject of great interest. A key characteristic of human faces is the ability to induce emotion in the viewer, through expressed emotion or other more abstract constructs such as trustworthiness or attractiveness. In this thesis, five functional magnetic resonance imaging experiments that probe the neural systems underpinning the perception of such emotive characteristics are described. I show that perception of emotional expression and identity are doubly dissociable with fusiform cortex encoding identity and superior temporal sulcus (STS) encoding expression. In subsequent experiments I explore the parameters under which distinct brain regions involved in emotional face perception engage, in particular addressing whether responses are automatic or dependent upon a particular task. The issue of whether distinct emotions are processed by different brain regions is considered and the basic stimulus property of spatial frequency is manipulated to address the idea of a subcortical visual pathway carrying emotional information. I describe two further experiments that address the more complex social constructs of attractiveness and trustworthiness, and demonstrate that broadly similar cortical circuitry is invoked when processing these attributes compared to basic facial emotions. Ultimately, a network of brain regions including amygdala, fusiform cortex, STS, and orbital and medial prefrontal cortex (OMPFC) is characterised as the substrate for emotional face perception. In general, I found that amygdala and fusiform responses to emotive faces are automatic, whereas STS and OMPFC responses show a greater degree of task-dependence. I interpret the amygdala response as an emotional labelling process, whereas fusiform enhancements to emotive faces probably reflect feedback from amygdala to modulate early face processing. STS responses indicate the encoding of specific facial expressions in this region and a wider role in intention detection. The response profile of the OMPFC is complex and I suggest multiple roles for this region in mediating an interaction between cognition and emotion.

Table of Contents

Table of Figures

Chapter 1: Introduction

The visual processing literature has a continuing debate as to whether, or to what degree, faces are "special" (Kanwisher, 2000; Tarr and Gauthier, 2000). By "special", a unique domain in the visual environment is meant, with a neural architecture devoted to facial processing. This thesis argues that faces represent special visual objects in a manner different to that previously considered. Specifically, it will be shown that faces convey emotive cues that demand and receive processing by the brain of a viewer. This does not imply that the emotional processing in response to faces is face-specific, nor is it concluded that faces are special relative to other visual objects, in that other visual objects are not tested. However, it is argued that these emotive aspects of face perception represent central and inescapable components of visual face processing, and serve to emphasise just why faces are such important visual stimuli.

The thesis begins with a brief overview of two complementary frameworks that have influenced the work described in later chapters. These are the fields of study of emotion and feelings (also known as "affective neuroscience"), and the more recently studied topic of social cognitive neuroscience. It continues with a review of the literature on the neural correlates of face processing, with emphasis on emotional aspects and data pertaining to humans. Finally, a brief summary of the approach and experiments presented in the later chapters is given.

The emotional brain – affective neuroscience

Emotions and associated feelings were extensively studied and written about by Charles Darwin (1872), William James (1884) and Sigmund Freud (1920) at the end of the 19th and start of the 20th Centuries, yet were subsequently largely neglected for a substantial period. The reasons for this are debated, but it seems probable that the study of emotions was brushed aside by the behaviourist movement in its eagerness to remove cognition and consciousness from the study of psychology. With the birth of cognitive psychology and subsequently cognitive neuroscience, the study of emotions has resumed with vigour. Intriguingly, some of the earliest debates are being revisited with the tools that modern neuroscience now offers.

There are various reasons why the study of emotion by neuroscientists and psychologists has begun to assume centre-stage in brain science. One is the rediscovery of the importance of a dissociation between emotion and feeling (LeDoux, 1996, 2000) which potentially offers a more tractable approach to the neuroscientific study of emotion (Damasio, 2001). Indeed Antonio Damasio has extended this distinction to formal definitions. Specifically emotion is defined as "a patterned collection of chemical and neural responses that is produced by the brain when it detects the presence of an emotionally competent stimulus...the responses are engendered automatically." This contrasts with feelings which "are the mental representation of the physiological changes that characterise emotions. Unlike emotions...feelings are indeed private, although no more subjective than any other aspect of the mind" (Damasio, 2001). This distinction is not universally accepted (see e.g. Panksepp, 2004); nevertheless it provides a useful heuristic for characterising studies in affective neuroscience.

This section begins with a brief review of studies pertaining to the neural substrates of emotion, and continues with a discussion of the rather more sparse literature describing studies which address feelings. A historical perspective is useful and is adopted where appropriate.

Emotion

The earliest attempts to describe emotional circuitry in the brain assigned emotional functions to *la grande lobe limbique* identified by Broca (1878) in the medial wall of each hemisphere of the brain (limbique means "edge"). Papez (1937) described a specific anatomical circuit in the medial aspect of the brain from anterior hippocampus, through the mammillary bodies, anterior thalamus, cingulate cortex back to posterior hippocampus. Much of this was predicated from observations of brain damage in rabid dogs which included the mammillary bodies. However, with the exception of cingulate cortex, few modern construals of emotion circuitry include the components of the Papez circuit. Around the same time, Klüver and Bucy were publishing reports of the effects of damage to the anterior temporal lobes (Klüver and Bucy, 1939, see below). Later schemas for emotional processing included the amygdala (later shown to be the key component in creating Klüver-Bucy syndrome by Weiskrantz, 1956). Paul MacClean (1949; 1970) proposed two complementary neuroanatomical theories of emotion that still influence thinking today. In early work (MacClean, 1949), he described the *limbic system*, bringing together the work mentioned above, and additionally including prefrontal cortex. Subsequently (MacClean, 1970), he suggested that three layers exist within the limbic system and brain, corresponding to three evolutionary stages. These

were the *reptilian brain*, which controlled reflex behaviours, *paleomammalian*, controlling learned emotion and *neomammalian*, whose function underpins complex cognition.

The theory of the triune brain remains influential. For example, the concept of cortical and subcortical routes to emotion (see below) might be construed as one contemporary correlate of this hierarchical view of emotion. In addition, one of the great achievements of the triune brain model was to emphasise the importance of interpreting brain function in an evolutionary context. However, the general concept of a unitary limbic system has been widely discounted by cognitive neuroscience and evolutionary psychology. Instead, a more modern conceptualisation of emotion is as a disparate set of modules, each evolved for specific purposes (Cosmides and Tooby, 2000). This is not to say that there is not overlap in their neuroanatomical substrates, but the underlying principle of organisation is presumed to be at a smaller scale than the grand schemas proposed by Papez and his antecedents. A well-characterised example might be the circuitry underlying fear conditioning (LeDoux, 2000). Extensive study has characterised circuits subserving this prime example of emotional learning to the level of brain regions, circuits, subnuclei, and molecular and genetic mechanisms. It is now clear that the same circuitry does not underpin other forms of emotional learning, such as instrumental conditioning (though there may be some overlap: for an example that shows commonalities and differences see O'Doherty et al., 2004). The brain regions now considered critical for generating emotions include specific brain stem nuclei, ventral striatum, hypothalamus, amygdala, orbital and medial prefrontal cortex (OMPFC), and anterior cingulate. The evidence for a role for each of these in emotions will be briefly discussed.

Brain stem nuclei

Specific brain stem nuclei involved in emotion include the ventral tegmental area (VTA) and its population of dopaminergic neurons (Schultz, 1998). These neurons respond to the receipt of unpredicted rewards and to the receipt of reliable predictors of reward. If a reliable predictor of reward is not followed by reward itself, firing rates in this population of cells decrease below baseline. These data (along with evidence from a number of studies using different techniques) strongly implicate the midbrain dopaminergic system in predicting rewards. It is less clear that this system is invoked in learning about punishing stimuli (Mirenowicz and Schultz, 1996; Ungless et al., 2004).

It has been suggested that the serotonin system is involved in longer-term alterations in emotion and mood. The evidence for this is largely from the success of serotonin-selective reuptake inhibitors (that boost serotonin levels in the brain by inhibiting its breakdown) in the treatment of depression (Blier and de Montigny, 1994; Fuller, 1995). In addition to nuclei controlling specific neurotransmitters, other brain stem nuclei control autonomic output and tone, which are suggested to be important in generating feelings (see below).

Ventral striatum

The ventral striatum, and particularly the nucleus accumbens, is amongst the primary targets of dopaminergic neurons from the VTA (Fallon and Moore, 1978). As part of the striatum it is thought to exercise control over action and movement and this accords with a view of the dopaminergic system in instrumental reward behaviours (Schultz,

1998; O'Doherty et al., 2004). Other parts of ventral striatum, or striatum generally, are implicated in other aspects of emotion. Caudate and ventral pallidum might play a role in the experience and perception of disgust (Calder et al., 2000, 2001), and other portions of striatum have been suggested to be selectively involved in anger (Calder et al., 2004).

Hypothalamus

A role for the hypothalamus in emotion was first demonstrated in the 1920s by Bard (1928). Cats with brain lesions above the level of the hypothalamus showed intact rage responses, those with lesions including this small region, however, demonstrated no rage to provocative stimuli. Subsequent research has demonstrated that subnuclei of the hypothalamus are involved in a range of emotion-related behaviours including control of appetite and thirst, the autonomic nervous system, and sexual behaviour (summarised in Rolls, 1999).

Amygdala

The amygdala is a structure in the medial temporal lobe of the mammalian brain, supposedly shaped like an almond ("amygdala" in Greek). It was first identified as a key structure in emotion when Weiskrantz (1956) demonstrated that many of the emotional deficits in monkeys with Klüver-Bucy syndrome (Klüver and Bucy, 1939) could be replicated by damage restricted to the amygdala. These features were subsumed under the heading "psychic blindness", meaning a lack of knowledge of the value of stimuli in the environment. Monkeys with lesions including amygdala showed

tameness, and a willingness to approach, mouth and mount objects that might normally have provoked fear responses. This role for the amygdala in fear-related processing has been consolidated by the large body of evidence for a role in fear-conditioning, mentioned above.

However, the amygdala is not only concerned with fear, as has repeatedly been demonstrated. Amygdala-lesioned rats or monkeys show impairments in instrumental conditioning to rewards (Hatfield et al., 1996; Parkinson et al., 2001), single neurons in monkey amygdala respond to the sight of rewards (Ono and Nishijo, 2000), and human amygdala responds to receipt of pleasant tastes (O'Doherty et al., 2001b), for example. Further evidence for a broad role of amygdala in emotion will be presented in this thesis, for example in Chapters 4 and 7.

Anatomically, the amygdala is highly interconnected with other regions thought to be important in generating emotions (hypothalamus, brain stem, ventral striatum and OMPFC) but in addition receives a variety of sensory input (Amaral et al., 1992). Highly processed information from all sensory modalities reaches the amygdala. In addition, it has been suggested that amygdala receives some pathways that transmit much coarser information (perhaps fewer than five synapses from the primary sensory transducers), at least in the visual and auditory domains (LeDoux, 1996). In the auditory domain this information has been shown to be functionally significant, in that conditioning to tones can proceed with only this pathway intact (LeDoux et al., 1984). The evidence for subcortical routes to amygdala in the visual domain will be discussed below.

The amygdala transmits signals to a variety of effector regions, including hypothalamus, brain stem nuclei and ventral striatum. In addition, it projects to all levels of visual cortex from V1 up to higher visual association areas in the temporal lobe (Amaral and Price, 1984; Amaral et al., 1992). These re-entrant connections are argued to have key functions in amplifying sensory signals of significance, as will be discussed in this thesis. Importantly, it has reciprocal connections with prefrontal cortex and OMPFC in particular.

According to some views, there is no unitary structure called the amygdala (Swanson and Petrovich, 1998). Instead, there is a loose collection of (up to thirteen) functionally disparate nuclei in the medial temporal lobe. To accommodate this view, some authors refer not to the amygdala, but to the "amygdaloid complex". This viewpoint is certainly provocative and is supported by a wealth of data highlighting the fact that different subnuclei of the amygdala have different afferent and efferent projections and functions. For the sake of brevity, in this thesis the amygdala is referred to as a unitary structure. It should be remembered, however, that it is a complex with definable sub-divisions.

There are a variety of models that purport to explain the functions of the human amygdala in emotion in one unitary manner. Two of the more important are the view of the amygdala as an ambiguity detector (Davis and Whalen, 2001) and as a "relevance" or "value" detector (Dolan, 2002; Sander et al., 2003). The former is supported largely by human neuroimaging data, particularly using emotional faces as stimuli (e.g. Adams et al., 2003; Kim et al., 2003). The latter view is somewhat broader, and encompasses aspects of the former. Both of these views emphasise a role for the amygdala in sensing important events or stimuli in the environment and enhancing sensory processing of

such stimuli. As noted above, however, the amygdala is a complex, and distinct sub-divisions may have distinct functions.

Orbital and medial prefrontal cortex

Damage to the orbital and medial prefrontal cortex (OMPFC) was shown to alter emotional and social behaviour by the famous railroad accident that befell Phineas Gage in the 19th Century (Harlow, 1848, 1868; Damasio, 1994; Damasio et al., 1994). Specifically, Phineas Gage suffered a lesion in his left OMPFC (with some damage to other prefrontal sectors in addition) and exhibited striking personality changes. He changed from being "an honest, reliable, deliberate person...to being childish, capricious and obstinate, showing poor judgement and being inconsiderate to others" (Harlow, 1868). The brief popularity of the prefrontal leucotomy in the 1950s also left many thousands of patients with severe personality changes, probably largely due to the resulting disconnection between OMPFC and the rest of the brain. Careful case studies of single patients have revealed similar effects (Eslinger and Damasio, 1985; Damasio, 1994) and recent studies of patients with discrete OMPFC lesions have demonstrated impairment in emotion recognition, and altered emotional behaviour (Hornak et al., 1996, 2003).

Techniques for studying the intact OMPFC such as single cell recordings in macaque monkeys and functional neuroimaging in humans have revealed activity in OMPFC that is related to the reward value of stimuli. For example, single cells in OMPFC respond to pleasant or unpleasant tastes (Critchley and Rolls, 1996a), odours (Critchley and Rolls, 1996c) or visual stimuli associated with rewards (Rolls et al., 1996). Some of

these responses are flexible, dependent on the organism's current motivational state – if sated on one reward type, responses to stimuli associated with that reward decrease (Critchley and Rolls, 1996b). Similar phenomena have been demonstrated in humans using fMRI – pleasant and unpleasant touch (Francis et al., 1999), taste (O'Doherty et al., 2001b), smells (Gottfried et al., 2002) and visual stimuli (O'Doherty et al., 2003b) have been shown to be represented in OMPFC as well as non-primary reinforcers, such as monetary reward or punishment (O'Doherty et al., 2001a). Specific representations based upon current reward value are also evident (Gottfried et al., 2003).

As with amygdala, it should be noted that the OMPFC is probably not equipotential across its entire spatial extent. Firstly, the posterior portions probably represent taste and smell in a manner akin to primary or secondary sensory cortex (Rolls, 1999). In addition, there is evidence that anterior segments possess dissociable functions. One theory that has evolved primarily from functional neuroimaging studies is that medial orbitofrontal cortex (OFC) represents rewarding stimuli, and lateral OFC punishing stimuli (O'Doherty et al., 2001a). This will be addressed in Chapter 7 of this thesis.

Anterior cingulate

The anterior cingulate is a bilateral gyrus that arches above the corpus callosum. Although included in Papez's anatomical schema, it is unclear on what grounds. However, it is certainly a key component of modern definitions of emotion systems. Recent research has identified a range of autonomic functions that are monitored or controlled by anterior cingulate, including cardiovascular responses (heart rate and blood pressure), electrodermal activity (Fredrikson et al., 1998; Critchley et al., 2000b,

2000c, 2001a; see also Critchley et al., 2003) and pupil dilation (Critchley, Tang, Glaser, Butterworth & Dolan, *unpublished observations*). Given the importance of autonomic activity in emotional responses, and possibly in generating feeling states (see below), the anterior cingulate seems a key locus for the interface between emotions and their cortical and subcortical generators.

It has been hypothesised that there are functional subdivisions to anterior cingulate, based upon meta-analyses of emotional and cognitive manipulations in functional neuroimaging studies, with emotional paradigms activating inferior, rostral regions and cognitive paradigms dorsal areas (Bush et al., 2000). The dorsal regions are similar, however, to those demonstrated to be invoked by autonomic control and monitoring, and it is possible that their activation in at least some supposedly cognitive paradigms reflects differential task difficulty and associated autonomic changes (Critchley, 2004). Other subdivisions of cingulate are more convincing: for example a portion of ventral cingulate underneath the genu of the corpus callosum (sub-genual cingulate) has been selectively implicated in low mood and depression in a number of studies (Drevets et al., 1997, 2002; Zald et al., 2002).

Emotional feelings

Relatively fewer studies have addressed the component of affect that is related to the conscious experience of emotions – emotional feelings. This is probably for a number of reasons. Firstly, it is only in humans that feelings can be studied convincingly. Secondly, methods based upon self-report went out of favour in the behaviourist movement in psychology, and were slow to be re-discovered. Thirdly, in techniques

such as functional neuroimaging that rely on comparisons between different conditions, it is not simple to find a comparison that highlights a feeling component alone.

One conceptual rediscovery that has advanced the study of the neuroanatomical correlates of feelings is the role of bodily responses and autonomic responses in generating feeling states (James, 1884; Damasio, 1994). This has been followed by a number of studies demonstrating regions of the brain responsible for autonomic control – anterior cingulate in particular, as mentioned above. In addition, the mapping of autonomic responses has highlighted insula cortex as a key component in what has become known as "interoceptive awareness" – the knowledge, accessible to consciousness, of bodily states (Craig, 2003). Insula activity has been shown to relate to blood pressure and heart rate changes, receipt of pain, and experience of disgust (Casey, 1999; Ploghaus et al., 1999; Critchley et al., 2000b; Wicker et al., 2003). In a study designed to test the role of insula in integration of awareness and feeling states, (Critchley et al., 2002) demonstrated enhanced insula activation to conditioned stimuli of which subjects were aware in subjects with intact autonomic nervous systems relative to a patient group with pure autonomic failure (PAF). The same patient group shows emotional change on a subjective questionnaire relative to Parkinsonian patients with similar levels of functional impairment (Critchley et al., 2001b), consistent with the proposal that autonomic responses are an important component of emotional feeling states (see also Heims et al., 2004).

Very few studies have involved asking subjects to experience emotions whilst in the scanner. In an early study, George and colleagues (1995) used happy, sad and neutral faces as well as recall of life-events to induce feelings in eleven female subjects. These

authors report widespread increases and decreases in regional CBF in relation to transient sadness and happiness. In particular, they emphasise frontal, temporal and cingulate as areas in which significant changes in blood flow are seen depending upon the emotion elicited. In a similar study by Lane et al. (1997b) subjects underwent ^{15}O PET scans whilst recalling happy, sad or disgusting personal events or watching film clips designed to elicit happy, sad or disgusting feelings. Neutral recall and film conditions were additionally included in the experimental design. In a subtraction analysis, whereby each emotion constituted one activation condition against its own neutral baseline, the authors report common and distinct patterns of activation for different emotions. However, in failing to contrast the emotional conditions themselves, the authors could make no strong inferences regarding differences between the distinct feeling states.

More recently, Damasio and colleagues (2000) have undertaken a large PET study addressing the functional neuroanatomy of feelings involving 39 subjects. A limitation of this study wherein distinct groups of subjects were required to experience sadness and happiness, fear and anger, sadness and anger, and fear and happiness, was a failure to counterbalance the pairings of emotions. Consequently, no subject was required to experience both sadness and fear, for example. As with the study of Lane and colleagues (1997b) the authors describe distinct patterns of activation in relation to different feeling states when contrasted with independent baseline conditions though there are no direct comparisons of activation conditions. An interesting finding was the increase in CBF in midbrain, pontine and hypothalamic areas regardless of which emotion experienced, supportive of the idea that emotions leading to feelings are associated with changes in autoregulatory (autonomic) functions controlled by such

areas (Damasio, 1994, 1999). In cortical regions, orbitofrontal, anterior and posterior cingulate and secondary somatosensory cortices all displayed rCBF increases in relation to the experience of feeling states. This pattern of activity is compatible with the authors' interpretation as reflecting mapping of first order changes in bodily states. However, in a PET study with limited temporal resolution, and in particular where feelings are self-generated, it is very difficult to separate cause and effect. Thus, it is possible that a component of the activity reported is actually associated with subjects' deliberate modulation of their own bodily states to produce the required feeling.

To conclude this brief review of data concerning feeling states, it seems clear that widespread cortical areas may provide their neural basis. The presence of subjective mood changes in patients with PAF (Critchley et al., 2001b) and OMPFC damage (Hornak et al., 1996, 2003) but not in patients with amygdala damage (Anderson and Phelps, 2002), is evidence of a neuroanatomical dissociation between regions involved in generating emotions, and those involved in generating or interpreting feelings.

The social brain – social cognitive neuroscience

There has recently been an explosion in interest in the neural underpinnings of social phenomena, or behaviour involving conspecifics. This enterprise is known as social cognitive neuroscience (SCN). Many recent reviews have exhaustively explored this discipline (see e.g. Adolphs, 1999), and the reader is directed to those more extensive reviews for a fuller treatment. The report of single neurons in the macaque IT cortex that responded specifically to hands (Gross et al., 1969) alerted researchers to the possibility that specific neural substrates for social processing existed and this report

was followed by many that described cells in higher visual regions with responses selective to faces (see e.g. Gross et al., 1972; Perrett et al., 1982; Desimone et al., 1984; Rolls, 1984). More discussion of these face cells follows below in the section concerning electrophysiological recordings during face processing. In the 1980s the possibility that autism could be characterised at least partially by a deficit in "theory-of-mind" – the ability to model conspecifics' cognitive processes – pointed to the existence at least one specific social module (Baron-Cohen et al., 1985). Subsequent neuroimaging studies in healthy individuals highlighted a network of regions (medial prefrontal cortex, temporal poles, and superior temporal sulcus [STS]) frequently active in making theory of mind judgements (Goel et al., 1995; Frith and Frith, 1999; Castelli et al., 2000; Gallagher et al., 2000, 2002; Gallagher and Frith, 2003).

Leslie Brothers wrote two classic reviews that might be considered the foundation of SCN (Brothers, 1989, 1990). She describes a neuroanatomical model of social cognition based primarily on evidence from single unit recordings and lesion studies in non-human primates. Specifically, the model highlights STS, amygdala and OFC as key regions for social cognition, along with the inferior temporal regions that had been shown to contain face cells. These areas, along with those subsequently described as being involved in theory of mind judgements remain the foundation of more recent models of SCN. A final addition to the brain areas hypothesised to mediate social cognition are the set of regions demonstrated to contain "mirror" neurons (particularly premotor areas, see e.g. Gallese et al., 2004; Keysers and Perrett, 2004). A brief description of evidence linking each of these areas to social cognition follows.

Amygdala

Besides those aspects of the Klüver-Bucy syndrome that might be considered social (e.g. tameness to conspecifics, hypersexuality), evidence for a role for amygdala in social cognition was found in a series of lesion experiments involving wild monkeys. Firstly, Rosvold and colleagues (1954) noted that dominance hierarchies in small groups of monkeys could be changed by lesions to the amygdala of members of the group. For example, following amygdalectomy, a previously middle ranking monkey became the clear dominant animal and the previously dominant monkey was an outcast. Subsequent studies indicated that adult animals re-released into the wild after bilateral amygdalectomy became socially isolated and died (Dicks et al., 1968). In addition, single cell recordings from monkey amygdala indicate that neurons in this region responded to complex social stimuli, including faces (Rolls, 1984; Leonard et al., 1985) but also to the sight of other monkeys moving, interacting or making expressive gestures (Brothers et al., 1990; Brothers and Ring, 1993).

Human patients with bilateral amygdala damage are extremely rare, but those that have been described frequently have deficits in perception of emotional expression in faces, which will be discussed in more detail below. Additionally, they are impaired on social judgements (such as trustworthiness or approachability) from faces (Adolphs et al., 1998) as will be discussed below and in Chapter 6.

Functional neuroimaging studies have implicated amygdala in a range of socially-oriented paradigms, largely involving faces (which will be discussed below). Additional evidence for a role of amygdala in social processing in humans has come

from studies addressing emotional bodily postures (Hadjikhani and de Gelder, 2003), perception of biological motion (Bonda et al., 1996), evaluative judgement of famous people (Cunningham et al., 2003) and theory of mind (Baron-Cohen et al., 1999; Fine et al., 2001; Siegal and Varley, 2002; Stone et al., 2003; Shaw et al., 2004).

The propensity of the amygdala to activate in social paradigms has led to the suggestion that in humans it has become more specialised for social processing than for emotion (Adolphs, 2003a). This is a provocative suggestion although it is clear that the human amygdala is responsive to primary emotional stimuli such as tastes and odours along with more complex social stimuli (Zald, 2003).

Superior temporal sulcus

Superior temporal sulcus (STS) was first considered a component of the social brain after the discovery of "face cells" in the anterior portions of this long sulcus in the rhesus monkey (e.g. Perrett et al., 1982; Desimone et al., 1984). The precise functions of STS in face perception will be discussed below and in Chapters 3, 4, 6 and 7, and in this section, I concentrate on social cognitive functions in STS not including face processing.

Some cells in monkey STS are responsive to certain head or body positions (Perrett et al., 1985, 1989; Jellema and Perrett, 2003; Eifuku et al., 2004). In addition, cells have been described in STS with responses selective to hand movements or even to specific hand movements (Perrett et al., 1989). Similarly, human hand actions are reported to activate STS, detected by PET (Grafton et al., 1996; Rizzolatti et al., 1996b). More

generally, biological motion, whether of the mouth, eyes or whole body, activates STS (Bonda et al., 1996; Puce et al., 1998; Puce and Perrett, 2003).

The role of STS in social cognition does not seem limited to face perception (see below) and processing of biological movement. A number of studies attempting to isolate neural signatures of processing related to theory of mind have also shown activity in the posterior portion of STS (Frith and Frith, 1999). These disparate functions which all activate STS (mouth movements, eye movements, biological motion and theory of mind judgements) have been unified by Frith and Frith (1999) as components of a system for detecting the intentionality of conspecfics. This is supported by evidence that the responses of some action-specific neurons in this region is modulated by the gaze direction of the monkey performing an action (Jellema et al., 2000), suggesting a sensitivity to the intentions of the agent. This interpretation of STS function will be considered in greater detail in Chapters 6 and 7.

Orbital and medial prefrontal cortex

Demonstrations of the involvement of OMPFC in social processing exist in a number of modalities. Single unit recordings in monkeys demonstrate responses to emotional faces (Critchley and Rolls, *unpublished observations*), consistent with lesion evidence that damage to this area causes impairments in emotion recognition (Hornak et al., 1996, 2003; Pixley and Bachevalier, 2002). Notably, deficits in emotion recognition extend into modalities beyond the visual domain (Hornak et al., 1996; Hornak et al., 2003), consistent with evidence that the OMPFC receives sensory information from all

modalities as well as other "limbic" regions (Carmichael and Price, 1995a, b; Rolls, 1999).

It has been noted that Phineas Gage and other patients with damage to OMPFC exhibit a variety of social deficits, consistent with a role for OMPFC in social cognition (Eslinger and Damasio, 1985; Damasio, 1994; Damasio et al., 1994; Hornak et al., 1996). Patients with damage to this region show abnormal skin conductance responses to social stimuli (Damasio et al., 1990b). Recent studies using the Wason card selection task (Wason and Johnson-Laird, 1972), wherein normal subjects perform better when a puzzle is framed in social terms than logically equivalent non-social terms (Cosmides, 1989; Cosmides and Tooby, 1992), have showed reduction of this enhancement in patients with brain lesions including OMPFC (Adolphs, 1999; Stone et al., 2002). Patients who suffer damage to this region early in childhood show impaired moral reasoning and behaviour (Anderson et al., 1999), consistent with evidence from neuroimaging studies of activation of this area when subjects engage in moral judgements (Moll et al., 2001, 2002a, 2002b). Additionally, it is suggested that ventral sectors of prefrontal cortex constitute a network for storage of social knowledge (Wood, 2003; Wood et al., 2003).

Regions implicated in theory of mind

A theory of mind (ToM) is the underpinning of cognitive predictions about the behaviour of conspecifics (Premack and Woodruff, 1978). The act of theorising about others' minds is also known as "mentalising" or adopting the "intentional stance" (Dennett, 1989). A number of neuroimaging studies have been undertaken to isolate the

neural substrate underlying ToM (Fletcher et al., 1995; Goel et al., 1995; Castelli et al., 2000; Gallagher et al., 2000; Vogeley et al., 2001) and three regions have consistently been activated in these studies (Frith and Frith, 1999; Gallagher and Frith, 2003). These are: a region of medial prefrontal cortex (mPFC) in the paracingulate gyrus, posterior STS, and the temporal poles. As Gallagher and Frith (2003) point out, only the paracingulate activation is specific to ToM. As discussed above, STS activates to a range of social stimuli (perhaps unified by the requirement of "intention detection"). Similarly the temporal poles are engaged by a variety of processes in which episodic memories are retrieved, including autobiographical memory (Fink et al., 1996), memory for emotional events (Dolan et al., 2000), and memory for familiar faces and voices (Nakamura et al., 2000, 2001).

Further evidence that the paracingulate is the key region in mentalising comes from two well-controlled studies in which the only difference between the mentalising and control conditions was the subjects' belief about whether they were playing a human or computer (McCabe et al., 2001; Gallagher et al., 2002). In both studies, paracingulate was activated but no significant activation detected in temporal poles or STS. Additionally, patients with Asperger's syndrome fail to activate paracingulate cortex when performing theory of mind tasks (Happe et al., 1996). More recent evidence is less clear, however, with one PET study demonstrating differences in activation between autistic subjects and controls in all three regions commonly implicated in mentalising (Castelli et al., 2002). Moreover, lesion studies have shown that extensive damage to medial frontal cortex is not necessarily accompanied by impairment on theory of mind tasks (Bird et al., 2004) whereas damage to posterior temporo-parietal junction sometimes is (Samson et al., 2004).

Mirror regions

"Mirror neurons" have been demonstrated in premotor areas in the macaque brain (Gallese et al., 1996; Rizzolatti et al., 1996a). These are cells which respond both when the monkey performs a specific action and when the animal observes that action (they have also been referred to as "monkey see-monkey do" neurons - Carey, 1996). It is postulated that mirror neurons offer a basis for the interpretation of the intentions of others, linking perception and action (Rizzolatti et al., 2001; Williams et al., 2001). Functional neuroimaging experiments have demonstrated that large portions of the human premotor and parietal systems are activated by action observation (Rizzolatti et al., 1996b) and that such areas are activated in a somatotopic manner, with the specific regions that are utilised in action also activated in observation (Buccino et al., 2001).

This general relationship between action and perception has been extended in a proposal that it also applies to the experience and observation of emotions (Adolphs et al., 2000; Decety and Chaminade, 2003; Gallese, 2003). Specifically, it is proposed that observation of an emotion in another, for example through facial expressions, is interpreted through simulation of the bodily ("somatic") state of the viewed individual on the viewer's cortical circuitry. This is supported by evidence from patients with brain lesions in primary and secondary somatosensory cortices, particularly on the right, who are impaired at emotion recognition from faces (Adolphs et al., 1996, 2000). Neuroimaging evidence for this proposal is lacking, a fact that will be discussed in Chapter 4 of this thesis. Evidence for aspects of this general proposition has come from two recent neuroimaging studies in which emotion and emotion observation were elicited in the same subjects in the realms of disgust (Wicker et al., 2003) and pain

(Singer et al., 2004b). In both studies a subset of the regions involved in experience of the emotion were activated by the perception or knowledge of someone else undergoing the same sensory stimulation.

Faces on the brain – neural systems for face processing

Patients with relatively selective deficits in recognition of faces were described by a number of researchers in the 19th Century (e.g. Wigan, 1844; Borelli, 1867; Quaglino, 1867; Hughlings Jackson, 1872; Quaglino et al., 2003). This suggested that the existence of a specific neural substrate for face processing. However, it was not until the late 20th Century and the development of techniques for studying the intact brain and characterising discrete lesions in the damaged brains of live patients that the neural underpinnings of face processing were thoroughly examined. As mentioned above, single cells in monkey IT cortex with responses selective to faces were observed in the 1970s and 1980s in a number of laboratories (Gross et al., 1972; Perrett et al., 1982; Desimone et al., 1984; Rolls, 1984). Around the same time, non-invasive imaging techniques that allowed observation of gross brain structure in living humans (e.g. computed tomography (CT) and MRI scans) began to become readily available. This allowed the anatomically-informed study of patients with specific cognitive deficits which had previously been limited to post-mortem examination. Amongst the cognitive deficits researchers wished to characterise was *prosopagnosia*, the inability to recognise faces (Bodamer, 1947). A key development in the study of face perception was the development of thorough cognitive models that described the subcomponents of what was clearly a multi-faceted process. The classic model is that of Bruce and Young

(1986), and there has been subsequent interest in characterising components of the Bruce and Young model neuropsychologically.

The Bruce and Young (1986) model posits separate modules for different aspects of face processing. At an initial "structural encoding" stage, physical characteristics of the face are decoded to match viewer-centred descriptions and expression-independent representations. Later stages of what is suggested to be a serial pathway represent higher aspects of identity processing, starting with "face recognition units" (FRUs), then "personal identity nodes" (PINs) and finally a name-generating module. Operating in parallel with this serial pathway and with one another are proposed to be a series of modules for "directed visual processing", speech analysis and importantly for our concerns here, facial expression analysis. The growing evidence for dissociable representations of different aspects of faces in the brain was reviewed by Haxby (2000; updated in Haxby et al., 2002) and anatomical localisations for some components of the Bruce and Young model were proposed. The suggestion was made that lateral temporal cortex encodes "changeable" aspects of faces (e.g. expression, eye gaze) whereas inferior temporal regions encode fixed aspects (e.g. identity). Evidence for these divisions, and for the division of expression analysis from identity processing in particular, will be discussed below, divided by experimental modality. I begin with discussion of data from behavioural experiments, and continue with neuropsychologically-oriented means of investigation: patients with discrete lesions, single cell recordings, event-related potentials and functional neuroimaging.

Behavioural data

Etcoff (1984) suggested on the basis of behavioural data that there are independent routes to expression and identity analysis. Specifically, she demonstrated in a Garner interference paradigm that subjects did not use information about identity that predicted expression to speed their decisions about emotional expressions. The absence of Garner interference suggests independent routes for processing, and the subsequent face processing model of Bruce and Young included parallel routes for these aspects of faces (Bruce and Young, 1986). More recent behavioural data (e.g. Schweinberger and Soukup, 1998; Schweinberger et al., 1999; Ganel and Goshen-Gottstein, 2004) has suggested less independence between identity and emotion processing, and will be discussed more in Chapter 3.

Investigations in brain-damaged patients

As mentioned above, a visual agnosia specific to faces has been described and is known as prosopagnosia. Prosopagnosia has been subdivided into *apperceptive* and *associative* (De Renzi et al., 1991). The former is a failure to perceive faces as a whole, and might represent a deficit in the structural encoding component of Bruce and Young's model. Associative prosopagnosia is an impairment in naming faces and might relate to damage to one of the modules involved in semantic associations with faces in Bruce and Young's model (FRUs, PINs, or name-generators).

There is little consensus on what specific brain damage causes prosopagnosia. Farah (1990) reports that 65% of patients with prosopagnosia had bilateral brain damage, a

relatively high proportion. Of those with unilateral lesions, there was a much higher incidence of right-sided damage (29%) than left (6%). In addition, there was not a very clear breakdown by lobe; although a large majority (>90%) had lesions that included occipital cortex, a significant proportion also had damage in temporal cortices (>65%). Other studies have also argued that inferior occipito-temporal damage (perhaps including the region of fusiform cortex shown to be responsive to faces in neuroimaging work - Kanwisher et al., 1997) is the most commonly damaged area in prosopagnosics (Damasio et al., 1990a). An important double dissociation exists between recognition of identity from faces and recognition of emotional expression. Prosopagnosics with spared expression recognition have been reported (Tranel et al., 1988) and identity recognition can be intact in patients impaired in emotion recognition from faces (Adolphs et al., 1999, 2000). This implies that different neural systems code for these different aspects of faces, consistent with their implementation in parallel pathways of the Bruce and Young model.

Two different types of study have produced different results with regard to patients with deficits in emotion recognition from faces. Group studies have found impairments in assessment of multiple different emotions, whereas single case studies have tended to emphasise impairments in recognition of one or a small number of emotion types. Examples of the former include Adolphs et al. (1996, 2000) and Hornak (1996), which suggest that brain damage in ventral frontal or right somatosensory cortices causes general in recognition of facial expressions of emotion. Single case studies include Adolphs et al. (1994, 1995, 1999) and Calder et al. (1996, 2000) and have suggested that amygdala damage causes relatively selective impairment in fear recognition and insula/basal ganglia damage impairment in recognition of disgust. A more recent

finding is an association of impairment in anger recognition and damage to ventral striatum (Calder et al., 2004). These findings will be discussed in greater detail in Chapter 4 of this thesis. Pertinent for the current discussion is the fact that the regions in which damage causes impairment in facial emotion recognition are dissociable from those causing prosopagnosia. Although right hemisphere is more frequently implicated than left in either deficit (see e.g. Etcoff, 1984; Damasio et al., 1990a; Farah, 1990; Adolphs et al., 1996, 2000; Kucharska-Pietura et al., 2003), more detailed anatomical investigations support the contention that identity and emotion recognition from faces are neuroanatomically dissociable.

One further study of impairment in face processing following brain damage of great relevance to this thesis is Adolphs et al. (1998). These authors demonstrated that regional brain damage was associated with a deficit in recognition of specific social traits (other than facial emotion) from faces. Specifically, it was shown that bilateral amygdala damage causes impairment in recognising approachability and trustworthiness in faces. This suggests that complex social traits may be encoded from faces by specific brain regions; a hypothesis explored further in Chapters 6 and 7. More recently, another study looking at patients with amygdala damage demonstrated severe impairment in recognition of social emotions from the face (Adolphs et al., 2002).

Electrophysiological recordings

Single cell studies in monkeys have demonstrated that inferotemporal (IT) cortex and superior temporal sulcus (STS) contain clusters of cells responsive to faces (Gross et al., 1972; Perrett et al., 1982; Desimone et al., 1984; Rolls, 1984). One study has explored

the possibility of a neuroanatomical dissociation between regions coding for identity and expression (Hasselmo et al., 1989), and suggested that while "face cells" in IT code for identity, those in STS code for expression (but see Tiberghien et al., 2003). More recent studies have also suggested a dissociation between these two areas with IT cells more invoked in identity representation across different views of the face and STS cells representing head direction across identities (Eifuku et al., 2004). Both of these suggestions are compatible with the distributed model of Haxby (2000), with its key point that lateral temporal regions code for changeable aspects of faces (e.g. head direction, expression) and inferior temporal regions for invariant aspects (e.g. identity).

With the exception of intracranial electrical recordings, relatively few of which have been reported, electrophysiological methods in humans have good temporal but poor spatial resolution. However, important information about the time course of and cognitive limits on face processing has been derived from event-related potential (ERP) data, and will be briefly reviewed here. The most familiar ERP component to researchers in face processing is the N170 (Bentin et al., 1996; George et al., 1996), a negative potential over occipital and temporal electrodes approximately 170ms after stimulus presentation. This effect is larger for faces than other visual objects, and is unaffected by familiarity of the face (Eimer, 2000), task (Mouchetant-Rostaing and Giard, 2003) or facial emotion (Eimer et al., 2003). Other ERP components are modulated by the presence of emotion in a face stimulus (Eimer and Holmes, 2002; Eimer et al., 2003; Holmes et al., 2003), or by directing visual attention to emotion or identity (Munte et al., 1998; Bobes et al., 2000). These data are supportive of the dissociation between structural encoding of faces independent of emotion or familiarity, expression analysis, and modules concerned with familiarity or identity hypothesised in

the Bruce and Young model (Schweinberger and Burton, 2003). One recent study suggested that activity of similar latency to the N170 in the fusiform gyrus was modulated by subjective "likeability" of the face (Pizzagalli et al., 2002). This is consistent with the idea that activity in fusiform cortex reflects affective judgements about faces, which will be discussed at greater length in subsequent chapters (see Chapters 4-6 and general discussion).

Intracranial recordings, which offer excellent temporal and spatial resolution, are rarely undertaken as they are performed only by necessity in clinical populations (typically to discern seizure focus in epilepsy). However, one set of published reports (Allison et al., 1999; McCarthy et al., 1999; Puce et al., 1999) uses this approach in a relatively large subject cohort (12 subjects) to characterise various aspects of face processing. Firstly, faces were distinguishable from objects by responses in ventral occipitotemporal cortex (approximating to fusiform gyrus) and lateral temporal regions (middle temporal gyrus, close to STS) with latencies of ~200ms (Allison et al., 1999). These lateral and ventral sites showed N200 response profiles that could be differentiated by stimulus characteristics such as eye-gaze direction (McCarthy et al., 1999), consistent with the idea that these two face areas subserve different functions in face processing (Haxby et al., 2000). Finally, early face-specific responses (N200) were relatively insensitive to top-down modulation by affective qualities of the stimulus, stimulus repetition, or semantic priming, whereas later responses (P350 and N700) were altered by such top-down processes (Puce et al., 1999).

While it is tempting to compare the intracranial recordings of N200s and face-selective fusiform responses (see below) to the face-specific scalp N170 (Bentin et al., 1996), it

should be noted that localisation of scalp ERP data is difficult because of the "inverse problem" – the fact that many different sources could account for the same patterns of activity recorded at the scalp. Even the most consistent ERP components such as the N170 are difficult to localise spatially, and debates continue as to whether the scalp N170 reflects activity in early occipital regions, fusiform, STS or some combination thereof (Schweinberger et al., 2002; Shibata et al., 2002; Henson et al., 2003; Itier and Taylor, 2004).

Functional neuroimaging

The availability of functional neuroimaging techniques since the 1990s has allowed leaps forward in the understanding of the neural underpinnings of face perception. The earliest studies were PET studies of simple face perception and demonstrated specific patterns of neural activity when subjects viewed faces relative to other objects (Sergent et al., 1992) or performed a face matching task rather than a spatial task (Haxby et al., 1991, 1994). Activation related to face processing included fusiform cortex on the inferior surface of the brain. Subsequent studies with fMRI (e.g. Puce et al., 1995, 1996) have demonstrated the reproducibility of this effect and led to this area being known as the *fusiform face area* ("FFA" - Kanwisher et al., 1997). The extent to which FFA activity reflects face-specific mechanisms rather than generalised systems for classification of visual objects of expertise (faces being objects of expertise for all healthy humans) is debated in the literature (Kanwisher, 2000; Tarr and Gauthier, 2000).

In addition to fusiform cortex, neuroimaging studies have tended to show activation in a number of other areas during face processing, including an area in lateral occipital cortex (sometimes called *occipital face area* [OFA]), superior temporal sulcus (STS) and less frequently, inferior frontal gyrus. The "distributed model" of face processing (Haxby et al., 2000) postulates separate roles for these areas, along the lines of the Bruce and Young (1986) model. Thus, the posterior occipital region is proposed to support structural encoding of the facial image, the fusiform invariant aspects of the face (such as identity) and the STS changeable aspects of the face (e.g. expression, eye gaze direction). Data from single unit recordings in monkeys described above (Hasselmo et al., 1989; Eifuku et al., 2004) broadly support this suggestion, as do imaging studies in which visual attention is directed to one or other aspect of the face (Sergent et al., 1994; Hoffman and Haxby, 2000). These data and this model will be discussed in greater detail in Chapter 3.

Neuroimaging studies concerning perception of facial expressions of emotion have been extensively reviewed (e.g. Adolphs, 2002a, b). One finding consistent across many studies is amygdala activation to fearful faces relative to neutral. Direct comparisons between different emotions been made in relatively few studies, but the consistency of amygdala activation to fear and its inconsistent activation to other emotions has led to the suggestion that different emotions are represented by different brain regions (Calder et al., 2001). This will be discussed at greater length in Chapter 4. In general, it is fair to say that few patterns of activity are consistent for emotions other than fear when comparing emotional faces to neutral, but see also the comments below about disgusted faces and insula.

The parameters under which fearful faces elicit amygdala responses have been extensively explored using neuroimaging. In the earliest studies, passive viewing (Breiter et al., 1996) or gender judgement (Morris et al., 1996) tasks were used to distract subjects from the blocked nature of the stimulus presentation. That differential amygdala responses between fearful and neutral faces were still detected implies that emotion need not be task-relevant to engage the amygdala (see below). Spatial attention to the stimuli is not necessary for amygdala activation (Vuilleumier et al., 2001), though larger responses are seen with directed spatial attention in other paradigms (Pessoa et al., 2002). Conscious awareness of faces appears also to be unnecessary – in masking paradigms in which subjects were prevented from consciously perceiving a fearful face by means of a rapid presentation and a subsequent neutral face, amygdala activation still obtained (Whalen et al., 1998; Anderson et al., 2003a). Similarly, fear-conditioned faces activated right amygdala even when masked (Morris et al., 1998b). Recent studies have demonstrated amygdala activity when conscious awareness of emotional faces is prevented by binocular suppression (Pasley et al., 2004; Williams et al., 2004). These findings have led to the suggestion that the amygdala might possess two sorts of inputs in the visual domain: a cortical route, receiving highly processed visual information from the anterior parts of the ventral visual stream, and a subcortical route from colliculo-pulvinar connections. As mentioned above, a similar dichotomy exists in the auditory domain in rat amygdala (LeDoux, 1996, 2000).

The idea of a visual subcortical route to amygdala has been tested in three ways. Firstly connectivity analyses of neuroimaging data reveal that functional amygdala connectivity to colliculus and pulvinar increases under conditions of fear-conditioned

faces when not consciously perceived (Morris et al., 1999; Pasley et al., 2004). Secondly, a blindsight patient, GY, shows amygdala responses to fearful or fear-conditioned faces when presented in his scotoma, and this amygdala response shows functional connectivity to posterior thalamus and colliculus (Morris et al., 2001b). Finally, recent data suggest that the amygdala is more sensitive to low spatial frequency fearful faces than high spatial frequency (Vuilleumier et al., 2003a), and any colliculo-pulvinar visual information would originate in the magnocellular pathway and thus be sensitive to low spatial frequencies (Tolhurst, 1975; see Skottun, 2000 for a discussion of transient/sustained and magnocellular/parvocellular distinction). These ideas will be discussed at greater length in Chapter 5.

One other facial emotion that has been relatively extensively studied is disgust. Evolutionarily, disgust (meaning "bad taste") is thought to relate to taste and smell (Rozin and Fallon, 1987; Rozin et al., 1994), although recent research suggests that it is strongly linked with disease sources and routes of infection (Curtis et al., 2004). This theory of disgust led to the hypothesis that regions of insula cortex involved in taste and flavour perception might also process disgust, and be invoked by perception of facial expressions of disgust. Anterior insula is frequently activated by disgust faces relative to neutral, though in some cases statistical comparisons to other facial emotions have not been made (e.g. Phillips et al., 1997), which precludes the conclusion that this is a specialisation for disgust. Where comparisons have been made, the predicted differences are not always clear-cut; e.g. Phillips et al. (1998) found significant differences in insula, but a later study (Phillips et al., 2004) did not. A recent study mentioned above included scanning components during both disgusting odour perception and viewing of faces of other subjects while they smelt disgusting odours

(Wicker et al., 2003). The same portion of anterior insula was activated in both conditions, providing strong evidence for an association between gustatory processing and disgust perception.

Perhaps the most consistent finding in studies which have compared emotional faces to neutral is that of fusiform activation (Breiter et al., 1996; Dolan et al., 1996; Morris et al., 1998a; Vuilleumier et al., 2001, 2003a; Pessoa et al., 2002; Surguladze et al., 2003), see also Chapters 4 and 5. This is not easily explained by the distributed model (Haxby et al., 2000), which postulates that facial characteristics such as emotion are processed in STS, rather than fusiform. Activation of fusiform by emotional expressions has been interpreted as reflecting re-entrant feedback from amygdala to enhance visual processing of the stimulus, but direct evidence for this is not strong (for exceptions see Morris et al., 1998a; Pessoa et al., 2002; Vuilleumier et al., 2004). This issue will be discussed in Chapters 3-5.

Another manipulation that has frequently been adopted in neuroimaging studies is comparing activation patterns in emotion-relevant tasks to tasks where the emotion in the face is irrelevant. There is some debate as to whether amygdala is activated in such comparisons, but since some studies report increased amygdala activity with emotional tasks (Baron-Cohen et al., 1999; Gorno-Tempini et al., 2001; Gur et al., 2002), others report decreases (Critchley et al., 2000a; Hariri et al., 2000; Hariri et al., 2003) and still others no differences (Narumoto et al., 2001), it is hard to know how to interpret these findings. One consistent finding is that STS is activated when subjects are asked to make direct emotional judgements relative to non-emotional judgements (Critchley et al., 2000a; Narumoto et al., 2001). This is consistent with the evidence that STS

contains neurons sensitive to facial expression (Hasselmo et al., 1989), and suggests that such neurons might be invoked in performing emotion-related tasks.

Relatively few neuroimaging studies have tackled the perception of more complex emotive traits from faces. There has been recent interest in the perception of race in faces, and much emphasis was placed on early results indicating that amygdala activity in Caucasian subjects to African-American faces correlated with implicit racist attitude on an independent standard test (Phelps et al., 2000). However, a recent lesion study failed to corroborate this finding (Phelps et al., 2003), and the functional significance of amygdala responses is thus unclear. A different study suggested that amygdala activity to same-race faces habituated with time whereas activity to other-race faces was sustained (Hart et al., 2000). A more recent study (Golby et al., 2001) suggested that fusiform cortex exhibited differential responses between same- and other-race faces, and that these differential responses correlated with memory superiority for same-race faces (a well-known psychological phenomenon, see e.g. Malpass and Kravitz, 1969; Brigham and Barkowitz, 1978; Brigham and Malpass, 1985; Chiroro and Valentine, 1995).

Attractiveness is a facial characteristic that has been extensively studied behaviourally, and recent neuroimaging studies have looked for brain systems that underpin perception of attractiveness in faces. Aharon and colleagues (2001) suggested a dissociation between beauty and attractiveness based upon selective responses in ventral striatum of male subjects to attractive female faces. O'Doherty and colleagues (2003b) found relatively few areas showing sex differences but did demonstrate responses to attractive faces in medial OFC and responses to unattractive faces in lateral OFC, concordant with

suggestions that these regions are dissociable in response to rewards and punishments (see the section "*The emotional brain – affective neuroscience: Orbital and medial prefrontal cortex*" above). Finally, one study exploring the interaction between eye-gaze direction and attractiveness indicated that dorsal thalamus and striatum showed increased correlation with attractiveness when eye gaze was directed at the viewer (Kampe et al., 2001). These results will be discussed at greater length in Chapter 7 of this thesis.

Eye-gaze direction itself is recognised as a powerful behavioural cue, and could be considered an emotional stimulus (Emery, 2000). Functional neuroimaging studies have found conflicting results in looking for neural expressions of gaze processing. Based upon the data from monkeys (e.g. Perrett et al., 1985) it might be expected that eye gaze direction in humans is processed primarily in STS. Although early results indicated that eye movements did indeed activate STS (Puce et al., 1998), subsequent data yielded from subjects looking at static gaze has not yielded very consistent results, with some indications that lateral temporal regions (including STS) are involved (Wicker et al., 1998; Calder et al., 2002), and other studies finding primarily amygdala and fusiform cortex (Kawashima et al., 1999; George et al., 2001). It is possible that these inconsistencies reflect the fact that regions that encode different eye-gaze directions (e.g. averted/direct) might be spatially contiguous (Perrett et al., 1985) below the resolution of conventional neuroimaging techniques. However the finding of amygdala activation to direct rather than averted gaze (Kawashima et al., 1999) and an association between amygdala-fusiform coupling and gaze direction (George et al., 2001) support the contention that gaze direction is an emotionally-relevant cue.

A final category of emotive face that has been studied is the faces of loved ones. Bartels and Zeki (2000) demonstrated that areas of the brain invoked by emotional experience such as insula and anterior cingulate are activated by viewing of pictures of partners compared to friends. These areas may be different from those activated by mothers observing pictures of their infants (Bartels and Zeki, 2004; Leibenluft et al., 2004; Nitschke et al., 2004), suggesting a dissociation between different forms of love.

Summary and introduction to the experiments in this thesis

A network integral to the processing of faces has been described (OFA, FFA, STS, amygdala), and a subset of these regions (STS, amygdala) along with a number of other brain areas (OMPFC, insula) described as being involved in emotional processing are likely candidates to subserve emotive aspects of face perception. The experiments described in this thesis were designed to probe this network and characterise the functional significance of each node. Chapter 2 constitutes an introduction to the methodology of fMRI (the technology used in all the experiments described in subsequent chapters). Chapter 3 describes an experiment designed to test for the hypothesised dissociation between invariant and changeable aspects of faces, specifically showing that fusiform regions process identity and STS processes emotional expression from faces. Chapter 4 reports a multifactorial experiment that attempts to dissociate brain regions involved in processing of specific emotional expressions rather than showing general emotion-sensitivity. Additionally, the effect of performing explicit emotion judgements on face stimuli is compared to a gender judgement task. Chapter 5 explores the underlying stimulus structure in fearful faces that might be driving emotional responses, specifically addressing the role of different

spatial frequency bands in providing input to regions sensitive to emotion in faces. Chapters 6 and 7 describe experiments designed to test emotive aspects of face perception beyond facial expression, addressing the more abstract constructs of trustworthiness and attractiveness.

Chapter 2: Methods

Introduction: Techniques for studying the active human brain

A range of techniques for studying activity in the intact human brain have been developed over the past century (Gazzaniga et al., 1998). The oldest of these, electroencephalography (EEG) permits excellent temporal resolution (in the order of 1-20ms), recording the electrical activity of the brain through the scalp. However, the spatial resolution offered is poor, on the order of magnitude of centimetres. Recent developments such as magnetoencephalography (MEG) offer similar temporal resolution, but also have limited spatial resolution (though this may be less limited than EEG methods).

Two technologies that offer improved spatial resolution, though poorer resolution in time, are based on detection of haemodynamic changes associated with neural activity. These are positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). fMRI offers a number of advantages over PET, specifically:

- fMRI is non-invasive (PET requires infusion of a radioactive agent)
- Improved spatial resolution (the resolution of PET is typically 10-16mm; resolution of fMRI can be as good as 1-3mm, though is more typically limited to 3-6mm)
- Improved temporal resolution (PET acquires an aggregate measure of blood flow over a 60-90s window; fMRI allows acquisition of single images in 1-3s, and has an effective temporal resolution of 3-10s)

- fMRI involves no exposure to ionising radiation, so can be used longitudinally or for within-subject controlled designs, such as pharmacological manipulations

fMRI does suffer some disadvantages by comparison to PET. Some of these are discussed in greater detail elsewhere in this thesis. They include: regional variability in time-courses of neuronal activation, the inability of current statistical models to account for potential regional variations in neurovascular coupling, and signal dropout and distortion in inferior temporal, inferior-medial prefrontal and brainstem regions.

This chapter begins with a brief description of the physics that underpins fMRI and continues with an introduction to the statistical analysis techniques used in subsequent chapters for fMRI time series analysis.

The physics of magnetic resonance imaging

How does (f)MRI work?

Magnetic resonance imaging (MRI) is a non-invasive technique that allows acquisition of images of human tissue *in vivo*. MRI relies on the phenomenon of nuclear magnetic resonance (NMR), reported by Bloch (1946) and Purcell (1946). In the experiments described in this thesis, one particular sort of MRI technique was used: functional MRI (fMRI). This approach describes the tracking of neural activity through MRI, of which a number of variants have been developed. Specifically, the experiments described rely on blood oxygenation level dependent (BOLD) fMRI, first described in humans by Kwong and colleagues (1992). In this section a brief explanation of the physical basis

for BOLD fMRI is given. More detailed descriptions can be found in Jezzard (2002), Buxton (2002) and Hornak (2003).

Spin

Spin is a fundamental quantum physical property of sub-atomic particles, like more familiar properties such as mass or charge. Unpaired neutrons, protons and electrons all possess spin of ±1/2. Pairs of like particles can cancel the detectable effects of one another's spin, so in MRI techniques, it is the spin of unpaired particles that dominate the signal. In practice this means that hydrogen atoms (with one unpaired electron and one unpaired proton) are particularly important, although other naturally abundant atoms such as carbon, nitrogen and sodium also have isotopes with net spin from unpaired nucleons. In particular, most of the signal detected by MRI in vivo comes from the hydrogen atoms that constitute water.

In the absence of a magnetic field, the spins of different nuclei within an object are randomly aligned, and there is no net direction of spin. Spins can themselves be viewed as magnetic dipoles (Figure 2.1), which means that when the object is placed in a magnetic field (B_0), the spins of the component protons align to the external field, either in the low energy state of S-N-S-N or the high energy state of S-S-N-N. A proton can flip from the low energy state to the high energy state by means of energy put into the system: this can take the form of absorption of a photon with energy exactly equal to the difference between states. The energy of a photon (E) is directly proportional to its frequency (ν), with Planck's constant (h) as the constant of proportion:

$$E = h\nu \qquad (1)$$

In addition, the frequency required is proportional to the strength of the magnetic field, with a scaling factor (γ) known as the gyromagnetic ratio:

$$\nu = \gamma B_0 \qquad (2)$$

Equation (2) can be substituted into (1) to calculate the energy required:

$$E = h \gamma B_0 \qquad (3)$$

For hydrogen, the gyromagnetic ratio is 42.6Mhz/T. This means that for typical field strengths used in fMRI experiments (1.5-3T) the frequency of the photon is the radio frequency waveband (64-128Mhz). In MRI, the frequency required to flip the spins of the protons is known as the Larmor frequency.

Net magnetisation

As described above, each nuclear spin can adopt either a high or low energy state. The ratio of high (N^-) to low (N^+) energy states is described by Boltzmann statistics:

$$N^-/N^+ = e^{-E/k\,t} \qquad (4)$$

t is the temperature, k is Boltzmann's constant (1.38×10^{-23} J/Kelvin) and E the energy required for transition between low and high energy states. At room or body temperature there are more protons in the low energy states.

Within a portion of an object in an external magnetic field, a group of spins will possess a net magnetisation. The net magnetisation will be in the direction of the external field (Figure 2.2a), which by convention is the Z-axis of a coordinate system (Figure 2.2b).

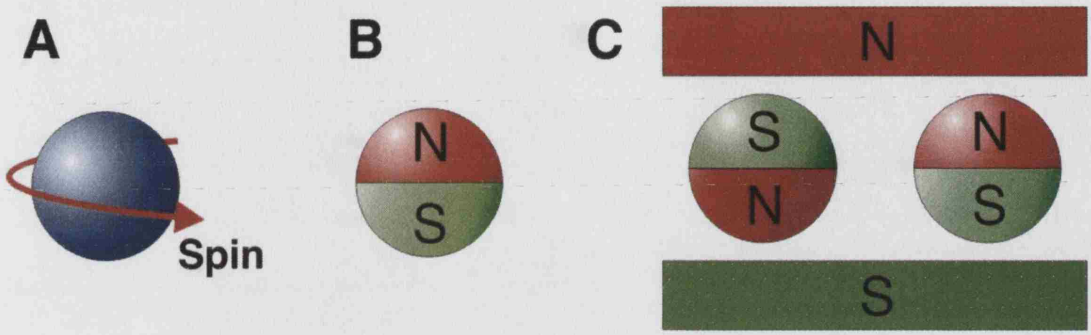

Figure 2.1: *Nuclear spins and magnetic dipoles*

a. Nuclei have the quantum property of *spin*

b. An alternative view of spin is as a magnetic dipole on the nucleus.

c. This means that when in an external magnetic field, there are low (left) and high (right) energy states of alignment between spin and the external field.

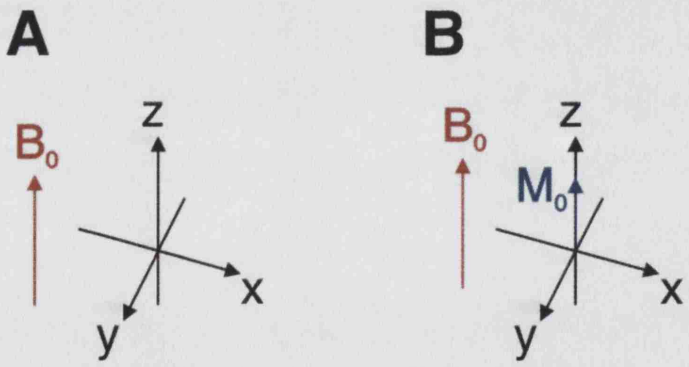

Figure 2.2: *Conventional coordinate system of MRI experiment*

a. By convention, the external field (B_0) is in the direction of the z-axis.

b. Net magnetisation (M_0) at equilibrium is in the direction of the z-axis.

This net magnetisation is proportional to N+-N-. At equilibrium, the net magnetisation is called the equilibrium magnetisation (M_0). Since this net magnetisation is in the direction of the B_0 field (the Z-axis), it is equal to the magnetisation in the Z-direction and M_0 is equal to M_z.

Radiofrequency magnetic fields and relaxation

As mentioned, even in an external magnetic field, nuclear spins can be perturbed by putting energy into the system in the form of a radiofrequency pulse. If the degree of energy put into the system is sufficient, M_z can be reduced to zero or even below zero. M_z will return to its equilibrium value with an exponential function, with time constant T_1. T_1 is also known as the spin lattice relaxation time. Since there is a distribution of rotation frequencies amongst particles, and only those molecules with rotation frequency equal to the Larmor frequency affect T_1, T_1 is altered by local temperature and viscosity through their effects on the distribution of rotation frequencies. As one contributor to tissue viscosity is fat content, imaging sequences sensitive to T_1 contrast can be used to delineate grey and white matter (they have differing fat content). In addition, the strength of the B_0 field affects T_1, because of the dependency of the Larmor frequency on B_0 (equation 2 above).

If the net magnetisation is not aligned with the B_0 field, the proton will rotate around the Z axis. This is known as precession, and the frequency of the rotation is equal to the Larmor frequency (Figure 2.3).

The net magnetisation can dephase across different portions of the sample (or *spin packets*), if the components experience different magnetic fields and rotate at their own Larmor frequency. The degree of dephasing depends upon the length of exposure to the perturbing field – increased exposure results in increased dephasing. Similar to the return to equilibrium in the Z plane, the return to equilibrium in the transverse plane is characterised by an exponential with time constant T_2 (analogous to T_1 in the Z axis). T_2 is known as the spin-spin relaxation time, and is always equal or less than T_1. Thus, the net magnetisation in the XY plane reaches zero before $M_z=M_0$.

Two independent factors combine to contribute to T_2.

1. "Inhomogeneous T_2 effect" resulting from variations in B_0
2. "Pure T_2 molecular effect" resulting from molecular interactions between adjacent molecules

The combined time constant of the two factors is T_2^* ("T_2 star"). It is T_2^* that fMRI sequences are typically optimised to measure.

Whilst the protons rotate in the XY plane, a nearby coil can detect the effects of the spins in the form of an induced current. The amplitude of this induced current is proportional to the density of protons detected within the sample. The coil used to detect effects can be the same coil used initially to excite the system with a radiofrequency pulse (an RF coil). As the system returns to equilibrium (relaxation), electromagnetic radiation is emitted as the protons return to a lower energy state. The signal received by the RF coil will gradually return to zero (Figure 2.4).

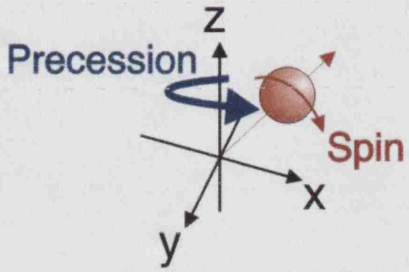

Figure 2.3: *Precession of particle with net magnetisation not aligned to B0 field*

Precession is around the z-axis

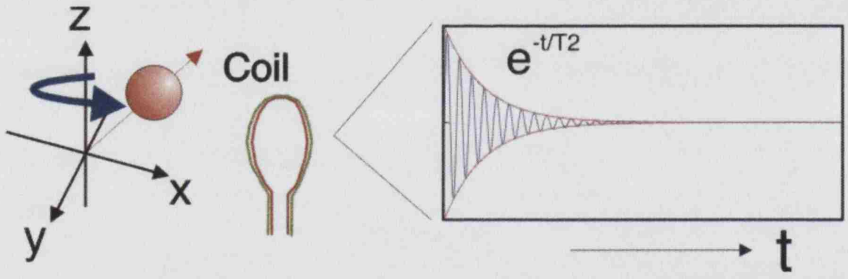

Figure 2.4: *Signal decay in simple MRI experiment*

The induced current in the coil is an AC signal which decays exponentially according to the time constant T_2. (The frequency in induced current is very high compared to that shown in the plot.)

Because different tissues have different T_2, one means of differentiating between tissue types using MRI can already be envisaged. By delaying the time at which we read the signal received by the RF coil, the sensitivity to different tissue types of the signal can be adjusted. For example with a B_0 field of 1.5T, white matter has $T_2 \approx 70$ms, grey matter has $T_2 \approx 90$ms and CSF has $T_2 \approx 400$ms. If the received signal is read after 100ms, signal from CSF will be high, having decayed relatively little, whereas signal from grey and white matter will be low, having decayed more rapidly.

The radiofrequency (RF) pulse used to perturb the system at equilibrium can be called a B_1 field and can be generated by passing an alternating current through a coil. A 90° pulse is one which causes the magnetisation vector to rotate 90° into the XY plane (a "90° flip angle").

MR contrast

The different types of MRI that are possible result from different combinations of RF pulse sequences, applications of magnetic field gradients and changes in sequence timing that sensitise the receive coil to different signals. In a simple 90° pulse and acquire sequence (as already described), if a second excitation pulse is applied shortly after a first excitation, the spins have not had time to return to equilibrium and the magnitude of the resulting signal from the second and subsequent pulses is reduced. If the second pulse is applied after a long gap (tens of seconds) the magnitude of the resulting signal can be as great as to the first. The gap between pulses is known as the repetition time, or TR. The speed of return to equilibrium is controlled by T_1, and different tissues have different T_1 characteristics (e.g. at 1.5T, white matter has

$T_1 \approx 700$ms, grey matter has $T_1 \approx 900$ms and CSF has $T_1 \approx 4000$ms). So a short TR (e.g. 600ms) will give higher signal from white and grey matter than CSF, as the system is closer towards equilibrium in the former tissue types after such a short TR.

Local inhomogeneities in the magnetic field cause protons to dephase. A "spin echo" pulse of 180° in the XY plane can refocus the spins. If the time after the 90° pulse at which the 180° pulse is applied is t, then at $2t$ the transverse magnetisation will peak again, with all spins in the same phase. $2t$ is called the "echo time", or TE, and such a signal is known as a "spin echo". Switching of magnetic gradients can produce a similar echo effect (known as a "gradient echo"; see below) and TE is a general term for the time after the RF pulse when the signal maximum is reached as a result of one of these echo-forming techniques. TR and TE are the adjustable factors in determining sensitivity to different tissue types in MRI. As described above, TR can be manipulated to adjust sensitivity to T_1 effects. If TE is much longer than T_2, then there will be little spin echo signal generated, and so little signal to detect. With TE shorter than T_2, there is little signal decay, so the signal is relatively insensitive to T_2. With TE similar to T_2, the signal detected will depend strongly on local T_2, and such a sequence is described as T_2-weighted.

Magnetic field gradients

Spatial localisation in MRI is performed by the use of relatively small magnetic fields that vary spatially to alter resonant frequencies across the space being imaged, after the techniques developed in the 1970s by Lauterbur (1973) and Mansfield and Grannell (1973). The MRI scanner has three gradient coils to generate these spatially varying

fields ("magnetic field gradients"). The three gradient coils produce fields in three orthogonal directions (x, y and z). The z-gradient coil produces a field that is zero in the centre of the magnet, more positive in one direction and more negative in the other. The size of the magnetic fields produced by the gradient coils is small compared to the B_0 field.

A key to acquiring MR images rapidly is the gradient echo. This is a phenomenon that relies on switching a field gradient on, then off, and switching it on in the reverse direction, and off again (Figure 2.5). Subsequent to the initial radiofrequency pulse, the protons precess coherently and the angle that each magnetisation vector makes in the XY plane is equal, as described above. When a field gradient is switched on, the magnetisation vectors of spins at different positions within the field begin to precess at different rates, because they are experiencing subtly different magnetic fields. The coherence of the transverse magnetisation is destroyed. Even once the gradient field is turned off although the rate of precession now equalises, the phase differences introduced by the field gradient remain. If a field gradient of opposite sign is now applied for equal time as the first had been, the phase differences are removed precisely, and the signals come back into phase, producing a "gradient echo".

Localisation

Three-dimensional MR imaging requires localisation of MR signals to different spatial locations. In any imaging process there is intrinsic uncertainty about the spatial source of a signal, and MRI is no exception. The unit of resolution in MRI is called a "voxel"

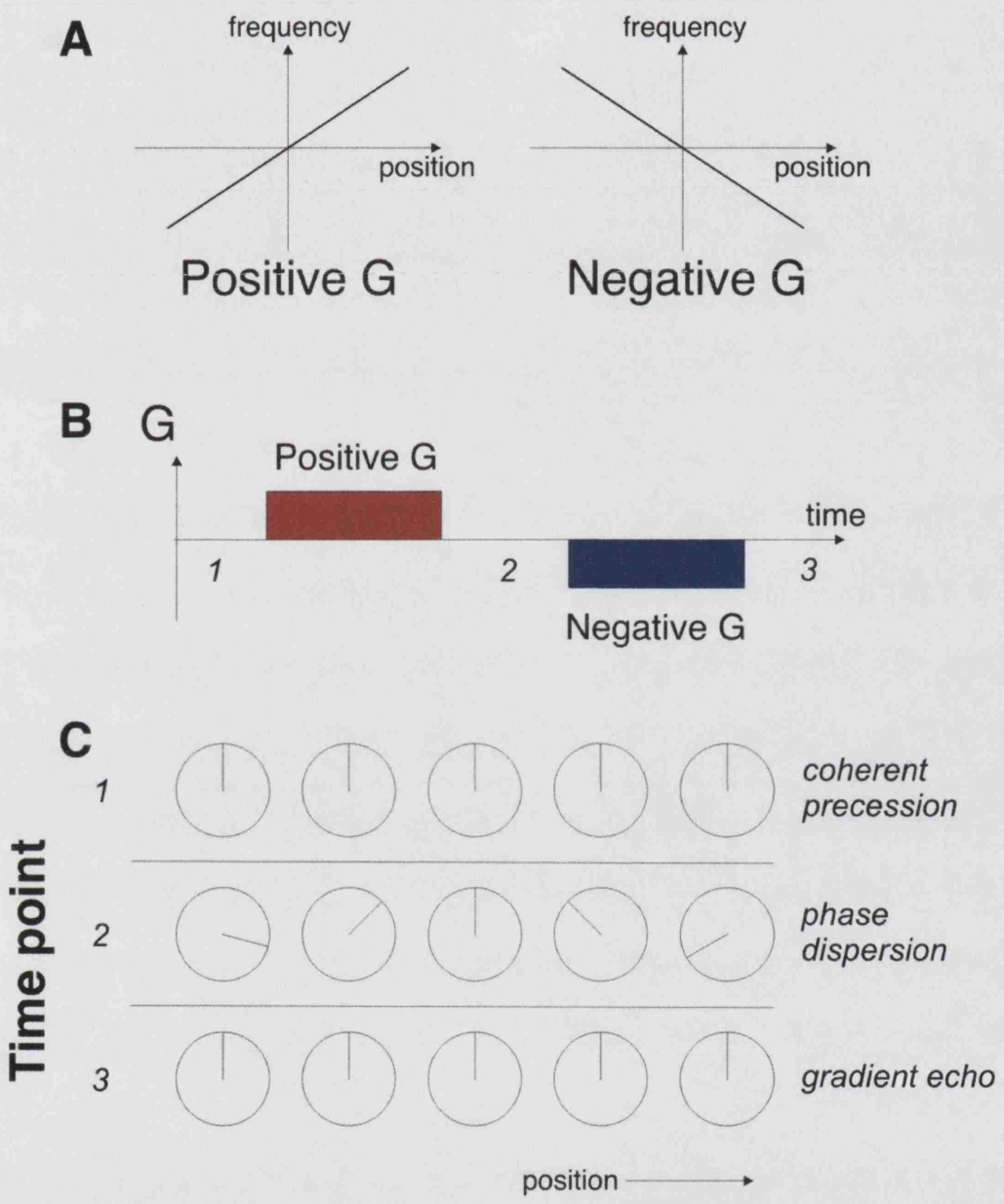

Figure 2.5: *Principle behind gradient echo formation* (After Buxton, 2002, figure 5.1)

a. The gradient varies over space such that the change in frequency of precession is related to spatial position.

b. The gradient is initially off (time point 1), then switched on and off (off at time point 2). The direction of the gradient is then reversed and kept on for the same amount of time as the positive phase until being switched off (3).

c. Initially (1), precession is coherent spatially. Subsequent to the application of the gradient (2) the phase of precession is dependent upon position within the gradient. After application of the reverse gradient (3), precession is again coherent (the "gradient echo").

(volumetric pixel), and is characterised by three numbers that describe the magnitude of uncertainty of localisation in each of the three dimensions x, y and z. Larger voxels mean that we are less certain of the source of a signal, but they may be more rapidly acquired.

Three means of localisation are most often used in MRI, each corresponding to one of the principal axes. These methods are: slice selection, frequency encoding and phase encoding, and correspond to localisation in the z, x and y axes respectively (though the directions of these axes do not necessarily have to relate to the use of z, x, and y terminology with respect to the B_0 field[1]). Each of these will be described in turn and are shown in Figure 2.6.

Slice selection

Slice selection limits the effects of the RF pulse to a single thin slice of the object being imaged. A magnetic gradient is turned on in the z-axis (the axis perpendicular to the desired slice) whilst the RF pulse is applied. The RF pulse itself has a highly specific range of frequencies, centred on ν_0. The effect of applying a field gradient in the z axis is to restrict the protons that will resonate with the RF frequency to a specific band perpendicular to the Z axis. Protons one side of this band resonate with too low a frequency, on the other side too high. The position of the selected slice can be adjusted by altering the frequency of the RF pulse, ν_0. The thickness of the slice can be adjusted by changing the steepness of the field gradient or by altering the range of frequencies in

[1] As it happens, for all the experiments described in this thesis, the two z axes were the same, since axial image acquisition was used, but this need not be the case.

the RF pulse. A typical slice thickness used in the studies described in this thesis is 2-3mm.

Frequency encoding

Frequency encoding allows spatial information about the source(s) of a signal to be encoded into the signal itself, as does phase encoding (see below). For frequency encoding, a negative field gradient pulse is applied in the x direction shortly after the RF pulse. This is followed by a positive x-axis gradient that spans the data collection window. The effect of this pair of pulses is to create a gradient echo midway through the positive pulse of the pair. Since data collection is occurring during this gradient echo, with the gradient field still turned on the precession frequency is varied linearly along the x-axis. The recorded frequency is a linear mixture of a range of frequencies and the component frequencies and corresponding spatial locations can be simply decoded from this mixture using standard Fourier analysis.

Phase encoding

Phase encoding relies on the application of a gradient field along the y-axis shortly after the RF pulse. With this gradient on, the transverse magnetisation precesses at different rates as a function of distance along the y-axis. Once the gradient is turned off, although the precession rate is no longer a function of position along the y-axis, the precessions along the axis are no longer in phase with one another (similar to time point 2 on Figure 2.5B and C above). The y-position of a given source is thus specified by a

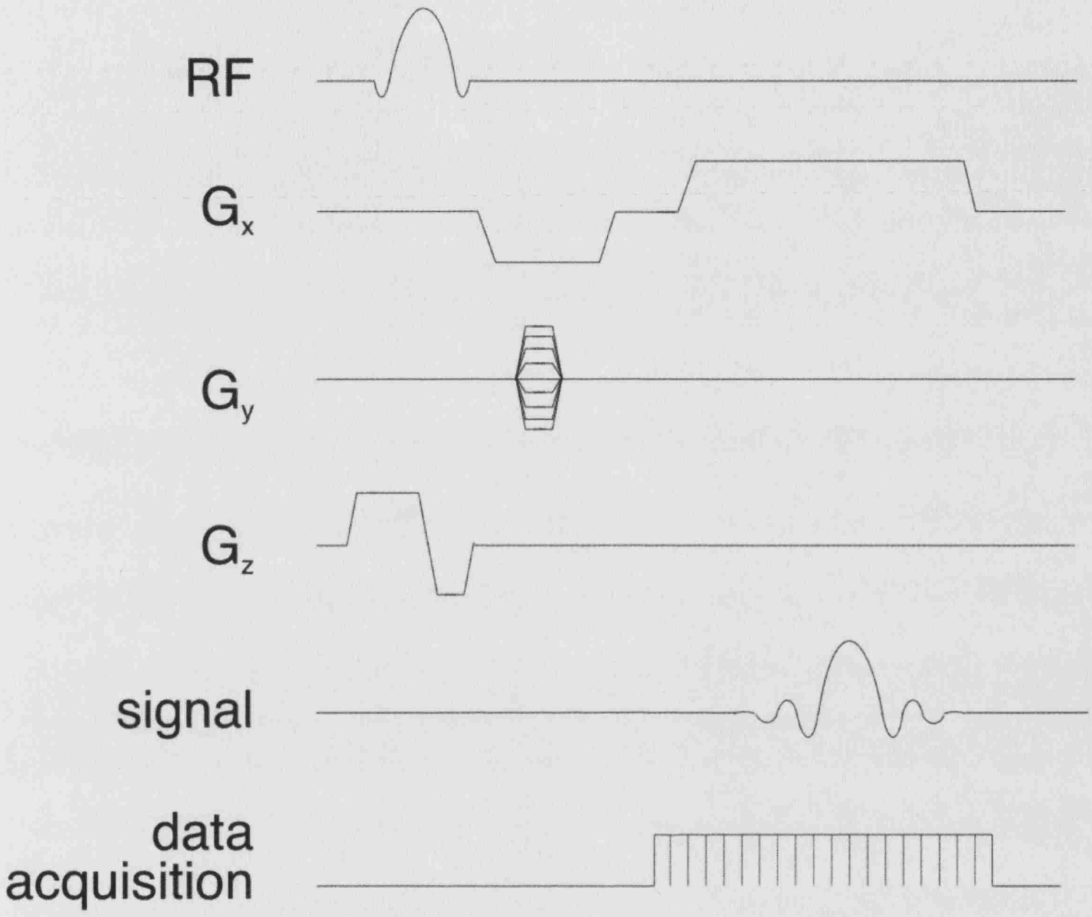

Figure 2.6: *Simple MRI sequence showing components important for spatial localisation* (After Buxton, 2002, figure 5.2)

RF is the initial excitatory RF pulse, G_x, G_y, and G_z are the magnetic field gradients for the x, y and z planes respectively, signal is the signal detected at the read-out coil, and data acquisition shows the time over which data is acquired. The initial RF pulse excites the sample in the B_0 field, whilst the G_z gradient is turned on to allow slice selection by varying the Larmor frequency across the sample in the z-axis. G_x is used to generate a gradient echo, such that the signal has a specific frequency depending upon the position along the x-axis. G_y can vary in amplitude to allow a specific phase to be encoded. The different values for G_y depicted allow different lines of phase space to be encoded, with one value (and hence one cycle through this sequence) for each phase encoding step.

particular phase of precession, and the x-position by a particular frequency of signal. Data acquisition with this sort of approach takes a number of iterations – this is because each component of phase space is sampled individually by altering the magnitude (or duration) of the y-gradient applied with each iteration (or "phase-encoding" step). This method of collecting MR images is used widely, but is unsuitable for dynamic imaging, wherein rapid image acquisition is necessary. The development in the late 1970s of echo planar imaging (EPI) made such rapid image acquisition possible.

Echo planar imaging

The innovation of echo planar imaging (EPI) is that a full image (of one slice) is acquired with a single RF excitation pulse (Mansfield, 1977). This is done by having a series of data acquisition points in conjunction with a series of rapid gradient switches in both the x- and the y-gradients – an example sequence is shown in Figure 2.7. The initial components of the sequence are standard: an RF pulse of appropriate frequency to match the slice-selective z-gradient, and subsequent application of gradients in x and y to permit frequency and phase encoding of the first sample point. Subsequently, the x-gradient is rapidly switched, which creates a series of gradient echoes for frequency encoding. With each switch of the x-gradient, the y gradient is pulsed on and off rapidly, to move through the phase of the precessing protons. With each gradient echo in x and pair of pulses in y come two data acquisition windows, which represent encoding of two lines of phase space (or two lines in the y-axis). The maximum resolution is limited by the physical difficulty of switching gradients on and off rapidly and the short duration of T_2^*. If the total window of data acquisition is much longer than T_2^*, then the signal will have significantly decayed by the latter sampling points.

The difficulty of switching the gradients rapidly limits sampling faster within the useful window. However, the window is adequate for acquisition of low resolution images, typically of 64 X 64 pixels per slice, or 3 X 3mm. The sequence continues with another slice selective z-gradient pulse and an RF pulse suitable to record a different slice of matter. The images can be reconstructed into three-dimensional volumes, though it must be remembered that different slices in the volume were acquired at different times. This can have implications for the modelling of neuroimaging data, which will be discussed below in the section titled *"Temporal realignment"*.

Blood oxygenation level dependent contrast

Rapid image acquisition with EPI is only useful to experimenters interested in functional brain activity if there is a metric of activity in the images. The typical metric used in fMRI and in all the studies described in this thesis is blood oxygenation level dependent (BOLD) contrast. It is known that increases in brain activity are associated with changes of blood flow to the active regions (see below). The key function of blood in this context is to carry oxygenated haemoglobin (oxyhaemoglobin) to active cells to allow oxygen exchange and permit efficient aerobic respiration. When fresh blood enters an arteriole or capillary, the ratio of oxy- to deoxyhaemoglobin rises. Deoxyhaemoglobin is paramagnetic due to its central iron particle with four unpaired electrons (Pauling and Coryell, 1936), and as a result, the magnetic susceptibility (which directly determines the magnetic field experienced by protons) of blood depends upon the ratio of oxy- to deoxyhaemoglobin (this is the blood oxygenation level dependency that gives BOLD its name).

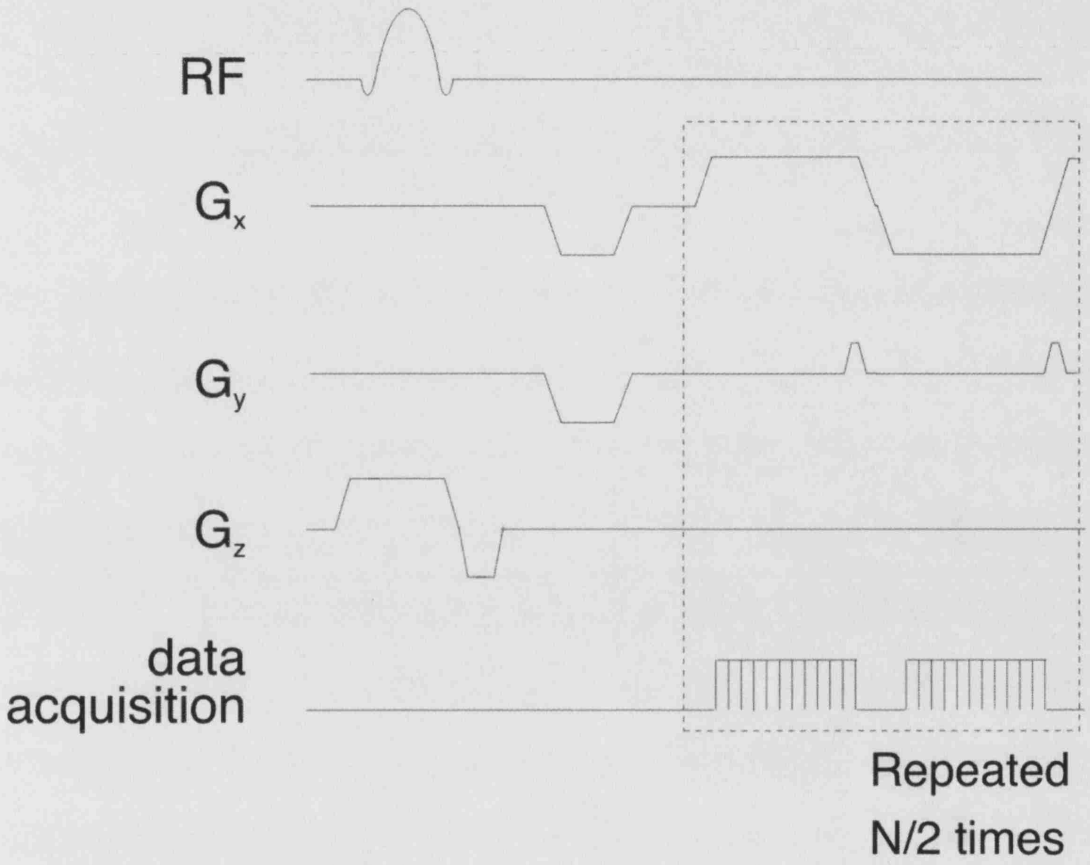

Figure 2.7: *Simplified MRI sequence for Echo-Planar Imaging (EPI)* (After Buxton, 2002, figure 11.7)

Here, repeated rapid switching of G_x with small increments of G_y allow sampling of phase and frequency space across the sample with a single RF excitation pulse. See text for details. (Labels as in Figure 2.6)

Protons in water molecules within the blood will experience local field inhomogeneities as a result of the levels of deoxyhaemoglobin. These inhomogeneities are detectable as a change in the T_2^* relaxation constant, with increased local fields causing shorter T_2^* and faster transverse relaxation. Thus, areas of low oxygenation and high deoxyhaemoglobin have increased inhomogeneities, lower T_2^* and will produce less MR signal for the same readout time. The feasibility of BOLD fMRI was demonstrated first in animal models (Ogawa et al., 1990) and subsequently (and near simultaneously by a number of research groups) in humans (Bandettini et al., 1992; Frahm et al., 1992; Kwong et al., 1992; Ogawa et al., 1992).

The sensitivity of MRI to BOLD is maximised by using a sequence sensitive to these microscopic changes in T_2^* that result from field variations introduced by the relative levels of deoxyhaemoglobin. To achieve this, a sequence with no spin echo (as this would remove field inhomogeneity effects which are a component of T_2^*) and moderate TE (tens of milliseconds) is used. One downside of the use of this long TE is increased sensitivity to signal dropouts due to field inhomogeneities caused by the presence of the head in the B_0 field. A shimming process is typically used to flatten out these inhomogeneities.

Because blood occupies only a small portion of the brain, BOLD fMRI depends upon detecting relatively small changes in T_2^*. At B_0 field strengths of 1.5-2T (used in the experiments described in this thesis), the signal change due to BOLD is typically of order of magnitude 0.5-2% of total signal. fMRI experiments involve sophisticated experimental design, image processing and analysis techniques in order to maximise

sensitivity to these small signals, which are described below in section *"Analysis of fMRI data"*.

The experiments described in this thesis were conducted on two different scanners at the Wellcome Department of Imaging Neuroscience, University College London. The experiments described in Chapters 4, 5, and 6 were conducted on a Siemens VISION (Siemens, Erlangen, Germany) system operating at 2T and those described in Chapters 3 and 7 were conducted on a Siemens Sonata 1.5T system. In all experiments described, T_1-weighted scans were acquired for detailed anatomical information and gradient echo, echo planar T_2^*-weighted images with BOLD contrast were acquired for fMRI. Resolution was 3 X 3 mm within plane of acquisition and slices were of 2mm thickness with a gap of approximately 1mm between slices. Thin slices have been demonstrated to improve signal quality in some of the regions of interest in this thesis such as amygdala (Chen et al., 2003; Robinson et al., 2004).

The neurophysiology of BOLD

That there is a relationship between neural activity and blood flow has been known for a long time (Raichle, 1998). Broca, more famous for studying patients with specific speech production deficits and their associated brain lesions, measured localised scalp temperature as subjects performed different mental tasks (Broca, 1879; reported in Cohen et al., 2004). An Italian physician, Angelo Mosso recorded the pulsations of human cortex in subjects with skull defects, even noting that the pulsations increased regionally with specific mental activities (Mosso, 1881). Roy and Sherrington conducted controlled animal experiments to explore the relationship between blood

flow and brain metabolism (Roy and Sherrington, 1890). In 1928 a neurosurgeon named John Fulton reported that by listening with a stethoscope at the back of the head of a patient with a damaged skull he could hear a louder sound when the patient's eyes were open, and louder still when the patient read from a text (Fulton, 1928). In other words, increased activity in visual cortex was resulting in increasing blood flow.

Despite over 100 years of study, the mechanism for this reflex is not clearly known, though it is hypothesised that local astrocytes respond to increased activity by dilating blood vessels and increasing flow and supply of oxyhaemoglobin (Magistretti and Pellerin, 1999). The time course of this response is somewhat stereotyped across brain regions and subjects (though to what degree is debated, see Aguirre et al., 1998; Miezin et al., 2000; Handwerker et al., 2004) and is described as triphasic with respect to BOLD recordings. In the first phase, increased activity in cortex leads to increased oxygen exchange with local blood vessels, causing a decrease in blood oxygenation (the "initial dip"). This has been observed in optical imaging experiments (Malonek and Grinvald, 1996) and in high field fMRI (Hu et al., 1997; Ugurbil et al., 1999). Vascular reflexes respond to the increased activity with increased local blood volume, and subsequently flow, bringing fresh oxygenated blood to the region and causing a rise in blood oxygenation and BOLD signal. This increase in oxygenation is typically an order of magnitude greater than the initial dip, and is the feature in the BOLD response that almost all fMRI studies detect (but see Ugurbil et al., 1999; Kim et al., 2000). Finally, relatively long after neural activity slows back down to baseline, blood oxygenation returns to baseline, with consequent reduction in BOLD signal. fMRI experimental designs and statistical analysis of fMRI data must take account of the fact that the haemodynamic changes to which BOLD is sensitive are considerably slower than the

neural activity which gives rise to the signal. Neural activity takes place on a time scale of tens to hundreds of milliseconds whereas the haemodynamic changes subsequent to that neural activity may peak after 4-6 seconds and last for as long as 20-30 seconds. The statistical methodologies for accounting for this will be described below in the section concerning analysis of fMRI data.

Recently there have been attempts to understand those aspects of neuronal activity to which BOLD activation relates (reviewed in Logothetis, 2002). Although one cross-species comparison suggests that for at least one region (MT/V5) there exists an approximately linear relationship between BOLD response in humans and single unit firing rates in macaque monkeys (Rees et al., 2000), more recent results imply a refinement of this position. Direct recording of single unit, multi-unit and BOLD activity in the same monkey brains suggests that it is the mass activity of neurons, characterised by local field potentials (LFP) that best predicts BOLD activation (Logothetis et al., 2001). This implies that instead of reflecting the neuronal output of a region, as a technique most sensitive to single unit spikes would do, BOLD fMRI best reflects the inputs to and local processing occurring within a portion of cortex.

Analysis of fMRI data

Introduction

The primary theoretical basis for experimental analysis in neuroimaging remains that proposed for cognitive psychology almost 150 years ago by Donders (1868, reprinted 1969), namely the subtraction approach. Donders worked with reaction times, but the approach can be generalised to any dependent variable, of which neuroimaging data is

one example. In the subtraction methodology, measured variables are proposed to result from underlying (invisible) cognitive processes. The difference between two dependent measurements is taken to reflect aspects of these underlying cognitive processes. For example, Donders found that discriminating a specific colour of light required approximately 50ms more than responding to the presence of a light alone. Response to the presence of a light represents a control state, and discriminating the specific colour an "activation" task. The subtraction of these conditions highlights a processing cost that reflects the enhanced cognitive demands of discriminating colour.

In the most basic implementation of neuroimaging data analysis, this basic subtractive approach remains the cornerstone. In neuroimaging, it is rare that one is simply interested in seeing the brain's response to a single condition and the most common approach is to compare an activation to a control condition. One reason for this is that carefully designed experiments can allow stronger characterisations of regional segregation: if one merely observed the activations resulting from observations of human faces, for example, much of the visual brain would be activated – reflecting the entire ventral and dorsal visual streams along with attentional and emotion systems, perhaps. Instead, a more sensible approach is to compare visual stimulation with faces to visual stimulation with other commonplace objects to isolate face-selective activations. The statistics underlying analysis of neuroimaging data remain tied to this subtractive approach.

Before statistical analysis of fMRI data (or "fMRI time series", since the data consist of a series of images in time) can proceed, a number of image processing steps are

undertaken to remove potential artefacts and to allow for appropriate statistical analysis. Broadly, these involve:

- ensuring the time series from one subject are aligned spatially, as subjects may move within the MRI scanner during the experiment (*realignment*)
- taking account of the fact that within a single three-dimensional image (or "volume"), different slices were acquired at different times (*slice-timing correction*)
- putting data from different subjects into a common anatomical space, to allow across subject (and across study) comparison and inference (*normalisation*)
- spatial smoothing of the time series to account for residual inter-subject variation in functional anatomy and to comply with requirements of the statistical analysis (*smoothing*)

In addition, a processing step that may sometimes be used involves putting images of different modalities from the same subject (e.g. T_1-weighted structural image and T_2^*-weighted functional images) into a common anatomical space, and this is known as *coregistration*. The logic and basic mathematics behind each of these steps will be described below in the section "*Pre-processing of fMRI time series*".

The statistical analysis of fMRI data used throughout this thesis is an approach known generically as Statistical Parametric Mapping, and the specific package utilised known as SPM (Wellcome Department of Imaging Neuroscience, London; http://www.fil.ion.ucl.ac.uk/spm). The approach adopted is *mass univariate*, i.e. each voxel is treated as an independent data series and a univariate (one dependent variable) statistical test is carried out at each voxel. Subsequent correction for the number of

statistical tests is then carried out, as conducting large numbers of tests renders one liable to false positives. The logic and mathematics underlying the statistical approach adopted in the experiments described in this thesis are described below in the section "*Statistical analysis of fMRI time series*".

Pre-processing of fMRI time series

Spatial realignment

A degree of head motion is inevitable in fMRI experiments given the common requirement for awake behaving human volunteers to remain still for up to one hour. This is minimised by use of light head restraint within the scanner during image acquisition, but residual movement persists. Spatial realignment estimates and corrects for the degree of head movement from scan to scan over an fMRI time series. Each volume is compared to a reference volume (in the experiments described in this thesis, the first volume of the time series was used as the reference) and parameters estimated for six affine rigid-body transformations. Affine transformations are manipulations in which parallel lines remain parallel, and rigid body transformations are merely translations and rotations. The six transformations estimated are x, y and z translations, and rotations in the three principal axes (or pitch, roll and yaw). The implementation of spatial realignment involves iteratively comparing transformations to minimise the mean squared difference between the current image and the reference, corresponding to a Gauss-Newton search (Friston et al., 1995a). A loosely similar calculation, without the iterative optimisation, was used in the more familiar domain of two-dimensional images to calculate visual similarity between face stimuli in Chapter 3 of this thesis (see the section "*Methods: Visual similarity between face pairs*").

Even after volumes are spatially realigned, subjects' movements during data acquisition might have introduced artefactual variance into the time series. This could either lead to increased Type I error through "activations" that are actually due to movement or Type II error by increasing the unexplained variance in the statistical model and reducing sensitivity to true effects. Sources of such variance include:

- Some MRI artefacts (e.g. Nyquist ghosts) do not behave as rigid bodies in head motion.
- Movement varies the timing relationship between the RF pulse and the selected slice, so more or less T_1 relaxation and consequently more or less signal may result from head motion
- Subject movement within volume acquisition may mean that each volume is not of identical size and shape (because different slices within volumes are acquired at different times), and consequently the rules of rigid body movement are not obeyed. Similarly, image distortions vary with head position in the scanner, and head movement may cause shape changes due to distortion of the volumes
- Interpolation in writing out realigned images can introduce errors

To reduce the possibility that the detected fMRI signals are due to movement-related artefact, the data are adjusted for residual movement effects. In the studies reported in this thesis, this adjustment was made by inclusion of the six estimated movement parameters in the design matrices for statistical analysis. Variance that correlated with head movement is therefore covaried out and does not affect statistical inference.

Temporal realignment

As mentioned above, the slices that constitute a volume are acquired sequentially, meaning that different slices are acquired at different times. In building statistical models to test fMRI data due account must be taken of this slice-timing problem. The degree to which this is a problem is dependent on the repetition time (TR) with which volumes are acquired. For medium-length TRs of 2.5-3.5s this effect can be dealt with using temporal realignment: interpolating the BOLD signal at each voxel in time to one reference point. For all the studies described in this thesis in which temporal realignment was used, the middle slice was picked as the reference and sinc interpolation used to adjust the data. Although for short TRs (such as that used in the experiment described in Chapter 3), this slice timing interpolation can also be applied, in practice it is not. This is because the slice-timing correction step is best done after the images are spatially realigned, but before normalisation. If slice-timing correction is not applied, the realignment parameters can be estimated and applied later, as part of the normalisation step. This limits the number of times that the images must be written out, with the associated image interpolation and loss of quality. With short TRs, the assumption is made that the temporal error introduced is negligible, a reasonable assumption with TRs of about 2s or below.

Spatial normalisation

Spatial normalisation of fMRI time series renders the volumes into a common anatomical space. This allows statistical inference about a group of subjects, as well as comparison of results between different experiments. The anatomical space adopted in

SPM is that of the template brain from the Montreal Neurological Institute, derived from 305 healthy volunteers, and it approximates to Talairach space (Talairach and Tournoux, 1988). The approach adopted in SPM (Friston et al., 1995a) consists of two components aimed at minimising differences between a source image (the image to be normalised) and an image in the anatomical space to be normalised to (the "template" image). These are:

- Parameters are estimated for 12 affine transformations: 3 translations (x, y, and z), 3 rotations, 3 shears and 3 zooms
- Parameters for a series of non-linear "warps" or deformations, estimated as the coefficients of a low spatial frequency cosine basis set

A Bayesian estimation framework is utilised to optimise the normalisation procedure. This attempts to find the set of deformations most likely given the data, and is implemented iteratively, using the mean squared difference between the image normalised by the current parameters and the template as the cost function to be minimised. Priors in the model are the likelihoods of a given warp, and "regularise" the procedure by preventing over-fitting to local features.

In the studies described in this thesis, the template image used for normalisation was the EPI template provided within the SPM package. This is a template derived from the averaged normalised EPIs from 13 different subjects. A representative image (a mean of all volumes for that subject for the studies in Chapters 4-7, and a whole brain EPI for the experiment in Chapter 3) was normalised to this template and the parameters derived from the normalisation procedures were then applied to each of the images in

the time series. In addition, where T_1-weighted structural images had been acquired, these were normalised using the parameters from the EPI normalisation after coregistration (see below) with the EPI from which the parameters had been estimated.

Spatial smoothing

Spatial smoothing is implemented for a number of reasons, some already mentioned. Briefly, these are:

- Central limit theorem implies that the distribution of errors resulting from smoothed data will be more normal, helping to ensure the validity of parametric tests.
- The matched filter theorem states that the smoothness of data should match the smoothness of an anticipated effect in that data for optimal sensitivity. The haemodynamic effects detectable by BOLD are probably larger than the typical voxel size used in fMRI experiments (of order 3-5mm).
- Inferences based upon Gaussian random field theory (see below) are predicated on the assumption that the error terms conform within reason to a lattice approximation of an underlying and smooth Gaussian field. For this assumption to be true, smoothness must be greater than voxel size.
- Smoothing data reduces inter-subject anatomical and functional-anatomical differences that remain after normalisation.

The normalised data in all the experiments reported in this thesis were smoothed in three dimensions with an 8mm full-width at half maximum (FWHM) Gaussian function.

Coregistration

Coregistration is used for within-subject conversion of images into the same anatomical space, and is typically (but not necessarily) used across different modalities. For example, in order to use the normalisation parameters estimated from EPI normalisation to normalise a T_1-weighted structural image, the starting space for the T_1 image must be the same as that for the EPI. Another reason to use co-registration is when restricted field-of-view (FOV) EPIs were acquired (such as the experiment in Chapter 3). Normalisation is improved by using as large a FOV as the template offers (typically the whole brain). Thus, whole brain EPIs are also acquired and used to estimate the normalisation parameters, and these parameters subsequently applied to the actual restricted FOV time series. Again, in order to apply parameters from one image to another, they must be in the same anatomical space, and coregistration is used to ensure this.

Because coregistration is designed for registering images across modalities, simple measures such as such as mean least squares differences between images cannot be used as cost-functions (the output measure that is to be minimised or maximised by the procedure). This is because there is no linear relationship between image intensities across modalities. Instead, co-registration in SPM works by maximising the mutual information between the two images. Mutual information is measured by summing over the two-dimensional joint histogram of the image intensities. Maximising the mutual information through the joint histogram maximises the overlap of the images' histograms, such that regions with similar intensities in the source image are best matched to regions with similar intensities in the target, without the stipulation that the

intensities for the source must be similar to those in the target. For example, assume grey matter has an intensity of 100 in the source and 120 in the target, and white matter an intensity of 110 in the source and 150 in the target and that we can ignore other tissue types. Maximum mutual information will then arise when voxels with intensities of 100 in the source match voxels with intensities of 120 in the target and those with intensities of 110 in the source match voxels with intensities of 150 in the target. The transformations allowed in coregistration are the same 12 affine transformations used for the linear component of normalisation. Again, the method proceeds iteratively until a maximum in mutual information is found.

Statistical analysis of fMRI time series

As mentioned above, the statistical analysis of fMRI data adopted for the experiments in this thesis is predicated on a mass univariate approach. In such an approach, a model is built of hypothesised effects, based upon the experimental design utilised. At each voxel in the brain, specific effects are examined by testing the fit of the model using a contrast pertaining to a possible effect (Friston et al., 1995b). Maps of t or F statistics are thus built up (statistical parametric maps, or SPMs), allowing a test of an experimental hypothesis against the null hypothesis of no effect. An inference about the effect can then be drawn after appropriate correction for the number of statistical tests performed. In SPM, parametric statistics are used to draw inference. This means that the statistics used have known distribution, and the probability of obtaining a particular statistical result is easily tested against this distribution.

Generalised linear model

The general linear model is the statistical machinery that underpins many common statistical tests, such as t-tests, correlation, multiple regression, analysis of variance (ANOVA) and analysis of covariance (ANCOVA). The general linear model is expressed in terms of a set of data (y), a set of estimated parameters (β), a set of designed effects (x) and an error term (ε). The errors must be independent from one another and identically distributed [*iid*, or more formally $\varepsilon \overset{iid}{\sim} N(0,\sigma^2)$]. Models with other distributions of errors are considered Generalised Linear Models, and strictly the abbreviation GLM is reserved for that more general case. Mathematically, each data point is expressed as the sum of the designed effects each multiplied by the associated parameter, plus the error term:

$$y_j = x_{j1}\beta_1 + \ldots + x_{jl}\beta_l + \ldots + x_{jL}\beta_L + \varepsilon_j \quad (5)$$

In this formulation, there are L explanatory effects and J observations. To express the entire model in one equation, matrix notation can be used. Here, the data, parameters and errors are column vectors and the designed effects a matrix:

$$Y = X\beta + \varepsilon \quad (6)$$

X is referred to as the "design matrix" and has L columns and J rows. β consists of a vector ($\beta = [\beta_1 \ldots \beta_l \ldots \beta_L]^T$), as does Y ($Y = [y_1 \ldots y_j \ldots y_J]^T$). The design matrix is a near complete description of the model, and factors not contained in X will be contained in the error term. The design matrix can be considered the quantification of our knowledge of an experimental design.

Typically there are more data points (rows in X, Y or ε) than explanatory variables (columns in X or rows in β), and the model constitutes a set of simultaneous equations

that cannot be solved directly. Instead, a method is required for finding the parameters that best explain or "fit" the data. The method adopted is ordinary least squares. Briefly, the least squares estimates are the parameter estimates ($\hat{\beta}$) which minimise the residuals. The residuals are the differences between the fitted data ($\hat{Y}$, which is equal to $X\hat{\beta}$) and the actual data Y. It can be shown that the least squares estimates, or maximum likelihood estimates are:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{7}$$

(X^TX) is only invertible if X is of full rank. As this is rarely the case, in practice the pseudoinverse of (X^TX) is used. If $(X^TX)^-$ denotes the pseudoinverse, then equation (7) can be re-written:

$$\hat{\beta} = (X^T X)^- X^T Y \tag{8}$$

which resolves to:

$$\hat{\beta} = X^- Y \tag{9}$$

Specific aspects of applying the GLM to fMRI data are considered below.

Using the GLM for fMRI time-series analysis: the design matrix

As described above, the GLM is the flexible basis underlying many common statistical tests. The different tests are represented by different design matrices, with common statistical tests utilising simple design matrices (see below). For fMRI time series analysis, the design matrix is constructed of experimental effects of interest and covariates of no interest, including (for example) the movement parameters estimated at the realignment step, a set of cosine functions to model slow drifts in the data and a constant term, to mean-correct the data. The experimental effects of interest are

typically an attempt to represent the experiment in terms of expectations about what BOLD signals would be expected, with individual columns, or *regressors* for each effect of interest. It is possible to use more than one regressor for each effect; for example, a number of differently-shaped regressors may be used if there is uncertainty about the shape of the expected response.

In event-related fMRI, which was the approach adopted for all experiments reported in this thesis, a brain region responding to a given event is assumed to show a BOLD profile with the characteristic haemodynamic response function (HRF) shape, time-locked to the stimulus onset (Friston et al., 1998a). As events are construed as very short transient phenomena, the event is initially represented as a stick or delta function. This is convolved with the HRF to produce a predicted response profile or regressor for each event type. The convolution assumes a linear time invariant (LTI) form, whereby events closer in time than the duration of the HRF simply sum linearly in BOLD activity. This assumption has been tested for distances between events (SOAs) down to about 1-2s, substantially less than the SOAs used in the experiments described in this thesis (Friston et al., 1998b). For SOAs of the magnitude used in the experiments described in this thesis, although there is some suggestion of reduced linearity in summation (Miezin et al., 2000), the reduction is relatively small (up to ~20%). In addition to the simple HRF form, other shapes can be convolved. Frequently the temporal derivative of the HRF is additionally used, as this can be useful in modelling HRFs that do not conform to the canonical shape (e.g. the peak is shifted in time).

Another extension that is commonly adopted is the modelling of parametric changes, to test predictions about the association of the haemodynamic or BOLD response with

some numerical variable associated with the stimulus (see e.g. Price et al., 1992). For example, the time in the experiment at which each stimulus occurred could be modelled in order to explore changes in signal over experimental time representing learning or response habituation. Alternatively, the stimulus itself may have some dimension that is captured by a parametric variable. For example in the experiments reported in Chapters 6 and 7 of this thesis, the stimuli were characterised by variation on facial characteristics such as trustworthiness or attractiveness. The relationship between these parametric variables of the stimulus and the brain's response were modelled by the inclusion of parametric factors in the design matrix. This is implemented in a very straightforward manner in SPM. The heights of the delta functions are simply modulated by the parametric variable of interest, before convolution with the HRF. In a simple extension, if one wants to explore non-linear relationships between the parametric variable and the haemodynamic or BOLD response (as in Chapter 7), the height of the delta functions can be modulated by a polynomial expansion of the parametric variable (Buchel et al., 1996, 1998a). Normally, the regressor of unmodulated height is included in the design matrix along with the linear and any other higher order expansions.

After construction of an appropriate set of regressors for each event type, the design matrix is completed by the inclusion of a set of covariates of no interest. For all the experiments in this thesis, these included the parameters describing the rigid body movements of the head estimated in the realignment step, a constant term for each session, and a set of cosine regressors modelling long term drifts in the data (which acts as high pass filter). Reasons for the inclusion of movement parameters were given above in the section "*Spatial realignment*". The constant term is included to mean

correct the signal from each session that a subject underwent and allow changes in signal to be estimated against a common baseline. The data are high pass filtered because a great deal of noise is detectable in low frequencies in fMRI data (Zarahn et al., 1997). Causes of such noise include physiological sources, slow movements and scanner drift over the course of the experiment. Because the experiments undertaken are designed, the experimenter can ensure that interesting signal should not be present at frequencies that are filtered out.

One problem with using the general linear model for fMRI time series analysis is that the data are serially autocorelated. This renders the error term non-independent (the first 'i' in iid is violated). To account for this temporal filtering can be applied to decorrelate (or *whiten*) the error terms. If utilised, the model corresponds to a GLM and the ordinary least squares parameter estimates are *maximum likelihood* estimators. In the experiments described in this thesis, inference was made at the random or mixed effects level (see below). Only the parameter estimates themselves affect this level of inference, and not the error terms about those estimates. Prior to improved machinery to estimate autocorrelations, the parameter estimates themselves were not altered by adjusting for serial correlations. Thus in the experiments described in Chapters 4 and 6, which were conducted before this machinery of restricted maximum likelihood (ReML) estimation was available, no adjustment was made. In the analysis of subsequent experiments ReML estimation (Kiebel et al., 2003) was used in order to render the error terms white for improved parameter estimation.

Interrogation of fMRI models in the GLM: Contrasts

Once the model is estimated, it must be interrogated to draw statistical inference, or to provide data for higher levels of analysis, such as random effects for group level inference (see below). This can be undertaken using contrasts of parameter estimates, or *contrasts*. The parameter estimates represent the strength of regression for each column in the design matrix. Thus, to compare the height of the HRF in two different conditions, parameter estimates for the HRF regressor are contrasted. This gives the size of the difference between conditions, which constitutes the numerator of the t statistic. The denominator is formed by the standard error of the contrast of the parameter estimates, obtained from the residual error of the contrast (simply the variance of the residuals around the least squares fit). In some situations, several contrasts of parameter estimates are jointly interesting. Examples might include the use of several different basis functions to characterise the shape of the haemodynamic response, or to characterise the main effects and interactions of a multifactorial design where one factor has more than two levels (e.g. Chapter 4 of this thesis). In this case, F-contrasts can be used to produce a map of the F statistic (an $SPM_{[F]}$). Subsequent to estimation of the statistics at each voxel, the map can be thresholded to allow statistical inference and rejection of the null hypothesis of no activation or no difference between conditions at a given p-value.

The machinery of SPM writes out new images with any t contrast performed. One image created is the statistical parametric map itself, a map with t-scores at every voxel tested. The second image is map of the contrast of parameter estimates, the numerator of the t statistic, which is the estimate of the effect size for the given contrast at each

voxel. The reason for this will be apparent later (see the section "*Population level inference: Random effects analysis*" below).

Correcting inferences for multiple comparisons

The use of a mass univariate approach to neuroimaging is statistically problematic, as a huge number (of order of magnitude 10^4-10^5 for fMRI time series analysis) of statistical tests are being performed. If conventional statistical thresholds (e.g. $p<0.05$) were adopted, a large number of voxels would be detected as showing effects simply by chance, i.e. there would be high Type I error and false positives would be detected. Thus, correcting for the multiple comparisons that have been made is necessary. A conventional means of correcting for multiple comparisons is the Bonferroni method. Here, one simply divides the p-value at which the null hypothesis will be rejected by the number of comparisons made. This is unsuitable for neuroimaging data, whereby because of inherent and applied smoothness in the data, the tests performed are not independent of one another. To take account of this smoothness, SPM makes use of *random field theory* (RFT) to control for a truer number of independent statistical tests.

A description of RFT depends on a concept called the *Euler characteristic* (EC). The EC is a characterisation of the number of clusters above a given threshold in a smooth field of random data. Thus, an image (e.g. of t scores) can be thresholded at a certain value, and the number of clusters with values above the threshold counted to calculate the EC. The expected EC (E[*EC*]) corresponds approximately to the probability of finding a cluster above a given threshold in the image. If one knows the number of independent units in the image, the E[*EC*] can be calculated for any given threshold.

The number of independent units depends on the smoothness of the image, and is measured in *resels* (Resolution Elements - Worsley et al., 1992). The E[*EC*] is proportional to the number of resels, conforming to the intuition that the more independent statistical tests performed, the more the likelihood of finding suprathreshold clusters by chance. The number of resels is calculated based on the smoothness of the residuals after statistical analysis (Kiebel et al., 1999). To reach a p-value threshold equivalent to $p<0.05$ allowing the number of tests performed, the Z-score threshold can be calculated (given the number of resels) that gives E[*EC*]=0.05.

Very frequently the experimental hypothesis refers to a specific brain region or set of regions. In this situation, correcting the p-values for the search space of the entire brain (as is the default in SPM) is inappropriate and could lead to false negatives. Instead, a correction for multiple comparisons over a small volume (or *small volume correction* – SVC) is appropriate. Here a shape is defined for the region – this can be a sphere centred on coordinates chosen from a previous study or a mask defining the region. Inference can proceed using the same RFT principles as described above.

In this thesis, results are described if they pass the threshold $p<0.05$ corrected for whole brain volume. Exceptions are made in *a priori* regions of interest where it is indicated if a voxel passed a threshold of $p<0.001$ without correction for multiple comparisons ("uncorrected") and additionally indicated if it passed an appropriate SVC.

Population level inference: Random effects analysis

Historically, there have been two primary approaches to the level of inference made from analysis of fMRI data (Friston et al., 1999a). These are fixed and random (or "mixed") effects analyses. In fixed effects analysis, the statistical tests performed are at the single subject level. To suggest that the effect pertains to the wider population, replications across a small number of subjects are frequently convincing. This approach can often be seen as similar to the classical case study approach popular in clinical studies. The numerator of the statistical test is the size of the effect in the subject under test, and the denominator the variability in measuring that effect. Typically, high degrees of freedom obtain, as the fMRI time series might contain several hundred data points in time. An alternative perspective treats the single subject as a noisy exemplar from a group, with inference made primarily at the level of the group – this is random effects analysis (Holmes and Friston, 1998). (The "random" is a reference to the treatment of the subject as a random factor.) In random effects analysis the numerator of a t statistic is the average size of an effect across the subjects and the denominator the variability of the effect in that population. The degrees of freedom are normally quite low, being approximately equal to the number of subjects (one less than the number of subjects for a one-sample t-test). The advantage of this approach is that the inference is at the level of the population from which the subjects were drawn. A significant result in a random effects analysis allows the inference that the effect exists in the population as a whole.

A random effects analysis is conducted in two stages in SPM. First, each of the subjects is analysed individually. This involves implementing a GLM as described above and

testing the contrasts that will provide the estimates of effect size at the random effects level. The images of contrasts of parameter estimates ("con-images") generated as part of this analysis (see the section "*Interrogation of fMRI models in the GLM: Contrasts*" above) are entered into a GLM as the second stage of the random effects analysis (the "second level"). The design matrix at the random effects level is normally a very simple implementation of a traditional statistical test such as a one sample t-test, two sample t-test, correlation analysis or ANOVA.

In the experiments reported in this thesis, random effects results are reported. The simplest design matrix possible that asked the appropriate experimental question was always used. For most experimental questions, this was a one-sample t-test looking at whether an effect was significantly different from zero across the group of subjects under study. The design matrix for a one sample t-test is simply a column of ones, of length equal to the number of subjects. For more complex experimental questions, such as those involving factorial designs where one factor had more than two levels (see below), conjunction analysis (e.g. Chapters 4, 5), or the between-groups study reported in Chapter 7, more complex design matrices corresponding to ANOVA designs were necessary. An issue that can be problematic in ANOVA designs is that of *non-sphericity* or *inhomogeneity of variance* – the fact that different groups or conditions can possess different degrees of variability, violating some of the assumptions underlying the general linear model. A second form of nonsphericity might be introduced by the fact that knowing something about a subject's activation in condition 1 might allow predictions about that subject's activation in condition 2: there may be correlation between the repeated measures in the ANOVA. In making inferences, this must be taken into account.

In classical repeated measures ANOVA designs, the error covariance structure is estimated to allow appropriate adjustment of the degrees of freedom for the F-statistic to remain valid. In SPM, there is the advantage of a large number of observations: the different voxels in the brain. By pooling over those voxels that are activated in the effects of interest F-contrast at $p<0.001$ uncorrected, a precise estimation of voxel-wide correlation matrix and voxel-specific hyperparameters controlling the nonspherical variance components can proceed (Glaser and Friston, 2004). The appropriate statistic and effective degrees of freedom (formally equivalent to a Greenhouse-Geisser correction) can then be calculated as described in Worsley and Friston (1995).

Design of fMRI experiments

As already mentioned, fMRI can be construed as simply another dependent measure in psychological research, just as response times or accuracy have been used for many years. I will discuss some aspects of experimental design with regard to this approach, and other aspects (efficiency of experimental design) with reference to a signal processing perspective.

Types of experimental design

As commented at the start of this chapter, statistical inference in most fMRI experiments still proceeds based upon a subtractive approach, by comparison of BOLD activation in two (or more) conditions. The simplest sort of experiment consists simply of two conditions, typically an activation condition and a baseline. The baseline should

be as similar as possible to the activation condition in all regards except the process of interest. Examples are familiar from early studies with PET (Lueck et al., 1989; Petersen et al., 1990; Zeki et al., 1991; Sergent et al., 1992). As is known from the statistical and psychological literature, however, a more powerful form of inference can obtain from *factorial designs*, in which a number of experimental *factors* are manipulated to test all *levels* of one factor against all levels of all other factors. Here, a level is characterisation of an experimental factor whereas a factor is a supraordinate category; for example an experiment might have factors of "colour" and "motion", wherein colour could have levels of "present" or "absent" and motion levels "high", "low" or "absent". This simple design would be referred to as a 3 X 2 factorial design, since there are two factors, one of which has three levels and the other has two. More complex designs can be envisaged: for example, a factor of attentional set can be added. Continuing with the example, "attention to motion" or "attention to colour" could be two levels of another factor. The design would now be referred to as a 3 X 2 X 2 design.

Factorial designs are considered more powerful than simple comparisons because they reflect the complexity of natural phenomena (i.e. rarely will a single factor work in isolation in nature) and they gain efficiency for the experimenter by means of the multiple experiments implicit within a single design (Fisher, 1966). Statistically, the advantage of factorial designs is that they allow a test for an *interaction* between levels of different factors. Put simply, an interaction is a difference in a difference, or the result that the behaviour of one factor depends upon the level of a different factor in the design. This is a powerful tool for investigating neurocognitive architecture (Friston et al., 1996). For many of the experiments in this thesis (e.g. Chapters 4, 6 and 7)

interactions will be used to examine whether manipulations in subjects' task affects the brain's response to emotive characteristics in faces.

Another key design type has already been mentioned: *parametric* designs. Here one of the experimental factors has many levels that can best be construed as varying ordinally or parametrically rather than categorically. Examples are seen in Chapters 6 and 7, where faces were rated by subjects as more or less trustworthy or attractive respectively. One can then test for a linear or non-linear relationship between the parametric variable and the dependent measure. Such designs can be factorial in nature too: in both examples in this thesis, the parametric factor is tested alongside other categorical factors.

Conjunction designs were introduced by Price and Friston (1997). The conjunction allows inference about a set of subtractions. One tests whether all the activations in a series of subtractions are jointly significant. If the series of subtractions contain one cognitive component in common, a region that activates in all the subtractions is associated with the common component. An example of the use of conjunctions is seen in Chapter 4. Statistically, a conjunction is akin to a split t-test, and Gaussian RFT has been extended to account for conjunction analysis (Worsley and Friston, 2000).

Design of efficient fMRI experiments

One useful perspective on design of fMRI experiments is that of signal processing. The BOLD response is construed as a signal to be detected, and experiments should be designed to allow efficient detection of that signal. From this perspective, the HRF can

be conceived of as a temporal filter on the underlying experimental causes (Josephs and Henson, 1999). Because the haemodynamic response to neural events is sluggish and long-lasting, it acts as a low-pass filter on neural activity, with maximal power at ~0.04Hz (25s). This means that maximally efficient designs are those that have experimental on-off manipulations with periods of ~25s. This equates to a block design, wherein a series of like stimuli are repeated for a period and followed by a block of stimuli of a different type. While maximally efficient, such designs are not optimal psychologically for a number of reasons, and where possible randomised event-related designs are used. *Event-related* (Josephs et al., 1997) means that the level of analysis is at single events, rather than treating a series of like events as one block. Randomised means that the order of trials is not predictable to the subject but randomised or pseudo-randomised. Randomised event-related designs have a number of advantages over blocked designs (Josephs and Henson, 1999), including:

- Responses to a trial are not systematically influenced by previous trials nor confounded by cognitive set (Johnson et al., 1997). This might be particularly important in the context of experiments pertaining to emotion, given the extensive literature on dissociable neural correlates of anticipation and receipt of emotive stimuli (e.g. O'Doherty et al., 2002).
- Trials can be categorised *post hoc*, by subject's responses. Examples in this thesis include the parametric studies described in Chapters 6 and 7, where *a priori*, the stimuli could not have been blocked by type. A further example is in Chapter 5 where subjects classified ambiguous stimuli and the response classified the trial type.

- Event-related fMRI is more directly comparable to other psychological disciplines where there is no reason to block stimuli, e.g. event-related potentials or psychophysics.

Blocks have the advantage of increased efficiency and the fact that some manipulations are hard to achieve in event-related fashion. For example, it is hard to force task switches frequently, so task-related factors are best blocked. However, given the advantages of event-related designs listed above, event-related fMRI was used for all the experiments described in this thesis (with task manipulations blocked). Simulations show that the efficiency of event-related fMRI is maximised when stimuli occur close together in time, i.e. the *Stimulus Onset Asynchrony* (SOA) is short (Friston et al., 1999b). Two slight caveats are necessary:

1. Although shorter SOAs are better, there is a point at which the LTI model of BOLD responses breaks down (Friston et al., 1998b), and below which shorter SOAs are less efficient (Friston et al., 1999b).
2. If one is interested in characterising the actual shape of responses relative to no activation ("evoked responses") longer SOAs are more efficient.

Point 2 can be circumvented by the introduction of an extra event type into the design: the *null event* (Friston et al., 1999b). Null events are not detectable by the subject, and constitute a trial of equal length to the experimental trials wherein no stimulus occurs.

All experiments described in this thesis used event-related fMRI with short SOAs of ~3s. This is above the SOA at which non-linearities strongly affect the additivity of the BOLD response, and short enough to allow reasonably efficient detection of experimental effects.

Chapter 3: Dissociation between identity and expression in face perception

Introduction

Models of face perception propose a dissociation between the representation of identity and other aspects of human faces, for example emotional expression (Bruce and Young, 1986). The "distributed model" (Haxby et al., 2000) suggests an anatomical basis for discrete stages of face perception proposed in the formulation of Bruce and Young (1986). In this model, fusiform cortex – a region known to be activated during face perception (Sergent et al., 1992; Puce et al., 1995, 1996; Kanwisher et al., 1997) - represents the identity of a perceived face, whereas superior temporal sulcus (STS) represents "changeable aspects" of the face, such as eye gaze and facial expression. Single unit recordings in monkeys (Hasselmo et al., 1989) and studies of human patients with discrete brain lesions (Young et al., 1993; Adolphs et al., 1996) provide support for this model.

Functional imaging studies comparing familiar versus unfamiliar faces also support a role for fusiform cortex in representing facial identity (George et al., 1999; Gauthier et al., 2000b; Henson et al., 2000, 2003). However, psychological differences related to processing familiar and unfamiliar faces, such as increased attention to familiar faces (Wojciulik et al., 1998), potentially confound these findings. Support for dissociable roles of fusiform and STS comes from studies in which attention is directed to different aspects of face stimuli (Sergent et al., 1992, 1994; Hoffman and Haxby, 2000; Narumoto et al., 2001, see also Chapter 4). Nevertheless, it is difficult to draw definitive

conclusions about the nature of the representations in these regions using task manipulations, because differences in activation might reflect processing involved in directing visual attention to the specific aspects of the face required by each task, rather than representations of those face components themselves.

Functional imaging studies of emotional facial expression have reported data that seem inconsistent with the distributed model, showing enhanced fusiform activity to emotional compared to neutral faces (Breiter et al., 1996; Dolan et al., 1996; Morris et al., 1998a; Vuilleumier et al., 2001, 2003a; Pessoa et al., 2002; Surguladze et al., 2003; see also Chapters 4 and 5). If fusiform cortex is specialised for identity it is unclear why it should show an enhanced response to emotional faces. This effect on fusiform cortex has been attributed to modulatory effects from amygdala, reflecting enhanced attentional processing associated with emotive stimuli (Dolan, 2002), but direct evidence for this proposal is sparse (for exceptions, see Morris et al., 1998a; Pessoa et al., 2002). A recent lesion study involving patients with medial temporal damage including the amygdala provides the most compelling direct evidence (Vuilleumier et al., 2004).

fMRI adaptation (fMRI-A) is a technique used to infer regional specialisation with greater specificity than the subtractive methodology used in most imaging studies. The logic of fMRI-A, outlined previously (Grill-Spector and Malach, 2001; Naccache and Dehaene, 2001; Henson, 2003), can be summarised as follows: if a region contains sub-populations of neurons excited by distinct aspects of stimuli, then when two stimuli are shown sequentially in which one of these aspects is repeated, the firing of excited neurons will habituate resulting in decreased fMRI signal from that region (compared to

when that aspect is not repeated). The claim is made that fMRI-A has smaller effective spatial resolution than subtractive imaging (Naccache and Dehaene, 2001) based on a dataset concerning number processing. When testing between trials containing just fours and sixes no differential effect was detected in a parietal region proposed to be important in number processing. However, *repetition* of fours or sixes within a trial led to a detectable difference from trials with both different numbers. Similarly, in the current experiment, a direct comparison of two emotional expressions (e.g. fear versus anger) might not yield activation in areas proposed to code emotional expressions (because the neuronal populations responsive to fear and anger overlap within such regions), whereas repetition of anger might yield lower activation than trials containing both angry and fearful faces.

In this chapter, I report the use of an fMRI adaptation paradigm to test the hypothesis, derived from the distributed model (Haxby et al., 2000), that fusiform cortex shows adaptation when identity is repeated (relative to when it changed), and STS shows adaptation when emotional expression is repeated (relative to when it changed). Given the factorial design (Figure 3.1), the prediction was that fusiform cortex should show a significant main effect of identity repetition and STS would show a significant main effect of expression repetitions. It was not anticipated that either area would show a significant interaction, as the interaction in this experimental design examines for areas showing a co-dependence between identity and expression repetitions.

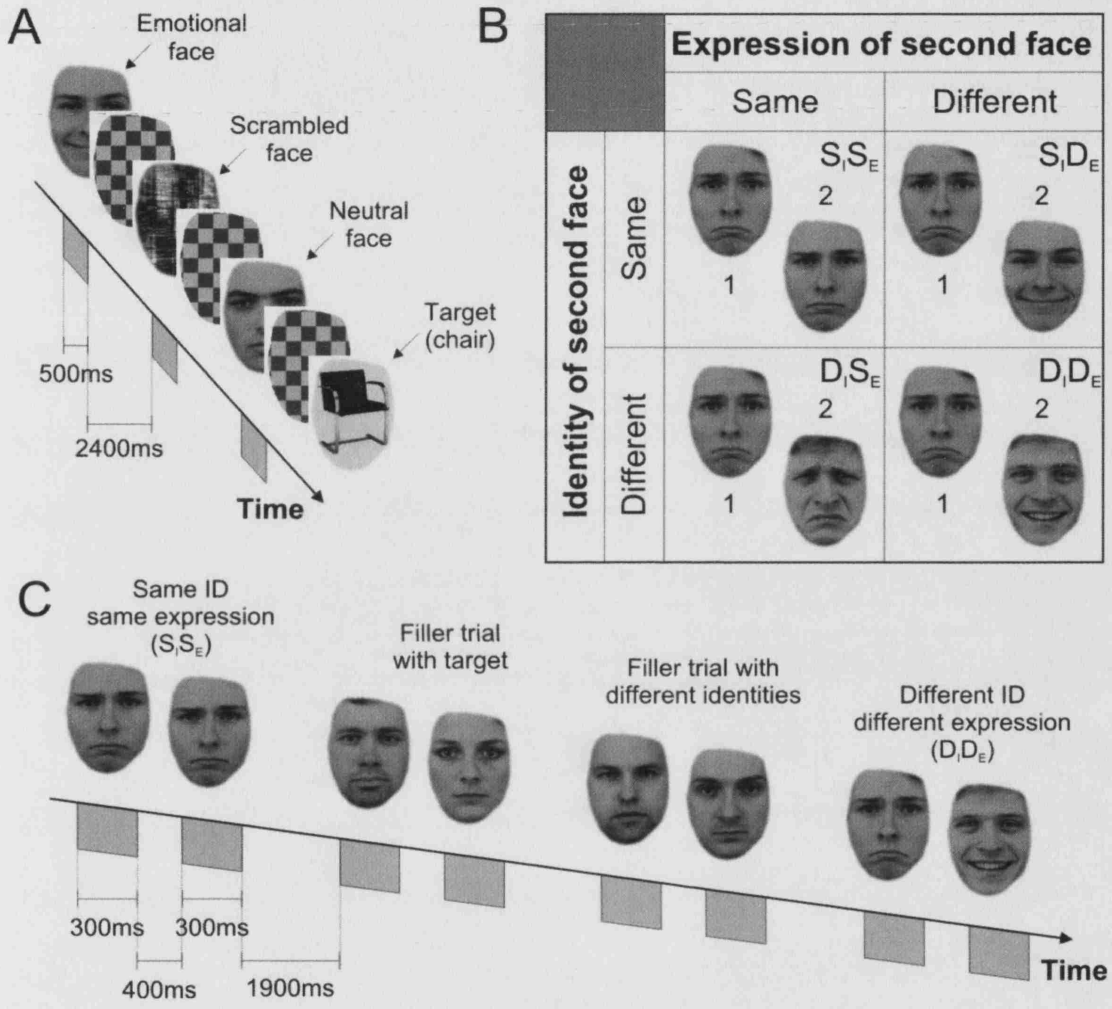

Figure 3.1: *Study design and example stimuli*

(Legend on following page)

Figure 3.1: *Study design and example stimuli*

a. In the localizer phase, subjects pressed a button when they saw chairs amongst a run of face and scrambled face stimuli. Stimuli were presented for 500ms with a 2,400ms gap during which a static checker-board within a face outline was presented.

b. 2 X 2 factorial design of adaptation phase. Trials in this phase consisted of a pair of faces which could display the same or different identity and (independently) the same or different emotion. Note that even $S_I S_E$ trials do not consist of identical stimuli, as the stimulus set offers two photographs of each individual displaying each emotion, taken on different days. Five different expressions (anger, disgust, fear, happiness and sadness) and five different identities were used. Note that the interaction in this factorial design does not pertain to regions showing specialisation for identity or expression processing; instead it tests for an adaptation to expression that depends on identity changes (or vice versa)

c. Example run and timing of four trials from the adaptation phase. Targets were rare (10%) trials containing female faces. Trials of interest were always separated by at least one filler trial consisting of neutral male faces which could either show the same identity or different identities. Similar to the localizer session (part a), but not shown for clarity, a face-outlined static checker-board separated all face stimuli.

Methods

I used the technique of fMRI-A, adopting a 2X2 factorial design (Figure 3.1B) to examine the neural basis for extraction of identity and expression from faces. Faces were presented in pairs where identity and expression in a second face could independently repeat or change with respect to the first face. Such immediate repetition induces robust fMRI adaptation (Kourtzi and Kanwisher, 2001; Chee et al., 2003; Epstein et al., 2003). Trials in which identity was repeated are referred to as "S_I" and I use "D_I" to indicate a change in identity across a face pair. Similarly, trials where expression is held constant or changed are labelled as "S_E" or "D_E" respectively. Thus, a trial in which identity was unchanged but expression varied is referred to as "S_ID_E".

Stimuli

The stimuli were a selection of five male faces from the KDEF database (Lundqvist and Litton, 1998). A major advantage of using this database is that it contains two exemplars of each expression for each identity. This allowed use of different images for the first and second faces in the adaptation session for all trial types, even S_IS_E. The five specific identities (males 13, 14, 16, 23 and 24) were selected on the basis of successful emotion recognition by nine subjects who took part in a pilot study judging the emotions expressed across the entire database. All face stimuli showed head direction and eye-gaze that were forward-facing towards the viewer. Five emotions were used in the study: anger, disgust, fear, happiness and sadness. One exemplar of each expression (series 'A') was nominated as the prime face and the second (series 'B') as the second face for the adaptation phase. Stimuli were converted to greyscale and equated for mean luminance in Matlab (The Mathworks, Natick, MA) and cropped

to a standardised outline in Photoshop (Adobe, San Jose, CA). Faces for the filler trials were neutral male faces from a variety of sources, and were prepared similarly to emotional faces, as were chairs and female faces (the targets in the localiser and adaptation phases respectively). 100 scrambled faces for the localiser session were derived from the 50 emotional faces and 50 neutral male faces used in that session by permuting the phase of each spatial frequency in the image while maintaining a constant power density spectrum, and then cropping to the same outline.

Subjects

Sixteen right-handed healthy subjects gave informed consent to take part in the study. Data were rejected from two of these subjects; one because of gross head movement during fMRI scanning, and the second due to an incidental structural abnormality that compromised normalisation of scans. Age range of included subjects was 18-29 years (mean=23), with seven males. Subjects had normal or corrected-to-normal vision.

fMRI experiment

There were three sessions to the fMRI component of the study. The first was a face localiser session, designed to familiarise participants with face stimuli and provide a generic map of face- and expression-responsive brain regions. The second and third were adaptation sessions and were split only for subject comfort. In the first (*localiser*) session, subjects' task was chair detection – they responded on an MRI-compatible button box whenever they saw a chair. The stimuli seen in this session were the 50 emotional faces from critical trials of the second and third (*adaptation*) sessions, 50

neutral male faces, 100 scrambled faces and 20 chairs. In the adaptation phase, the task was to press a response button when a female face appeared. In this way, trials of interest in all sessions were uncontaminated by motor response and repetition of faces was incidental to the subject's task. A number of different trial types occurred in the adaptation phase. The four trial types of interest consisted of a pair of emotional faces that could exhibit the same or different identities, crossed with the same or different emotional expression (Figure 3.1B). In addition, between any pair of trials of interest, one or two "filler trials" occurred to reduce the predictability of repetition. Three types of filler trials were used – "repeated fillers", in which the face pair consisted of two neutral faces of the same identity; "different fillers", consisting of a pair of neutral male faces of different identities; and "target trials" containing a neutral male and a neutral female face (which could be either the first or second face of the pair). In total, 80 repeated filler trials, 60 different filler trials and 20 target trials occurred in each of the two sessions. 25 trials of interest occurred for each of the four trial types in each session. Because there is only one way of combining the faces to produce $S_I S_E$ trials, but multiple ways of combining to produce the other trial types, the face pairs used for such trials were counterbalanced across subjects, with the overall constraint that each first face occurred once in each condition and each second face once in each condition within each session.

fMRI scanning

A Siemens 1.5T Sonata system (Siemens, Erlangen, Germany) was used to acquire BOLD contrast weighted echoplanar images (EPIs) for functional scans. Volumes, which consisted of 24 horizontal slices of 2mm thickness with a 1mm gap, were

acquired continuously every 2.16s. This sequence was sufficient to obtain coverage from above the corpus callosum to below the inferior temporal lobes, thus including all regions of interest in this study: fusiform, amygdala, STS, inferior frontal cortex, and orbitofrontal cortex (Figure 3.2). In-plane resolution was 3 X 3mm. The first six volumes were discarded to allow for T1 equilibration effects. Subsequent to functional scans, a T1 weighted structural image (1 X 1 X 1mm resolution) was acquired for coregistration and display of the functional data. Because of the difficulty of normalising limited field-of-view EPIs, whole brain EPIs were additionally acquired in each subject for improved normalisation.

Spatial pre-processing and data analysis

fMRI data were spatially preprocessed using SPM2 as described in the general methods chapter. Normalisation parameters were estimated using whole brain EPIs and applied to the restricted field-of-view images from the functional run.

Data analysis used SPM2, applying a mass univariate general linear model (GLM) as described in the general methods chapter. Included in the model were three regressors of no interest corresponding to the potential confound of similarity between face pairs (see below). Serial autocorrelations were modelled using an AR(1) process and the data were high pass filtered at 1/128Hz. Linear contrasts pertaining to the main effects and interaction of the factorial design were calculated. Consistent effects across subjects were tested using the resultant contrast images in one-sample t-tests (conforming to a "random effects" model). The model for the localiser session included separate regressors for the distinct facial expressions. No trends ($p<0.05$ uncorrected) to

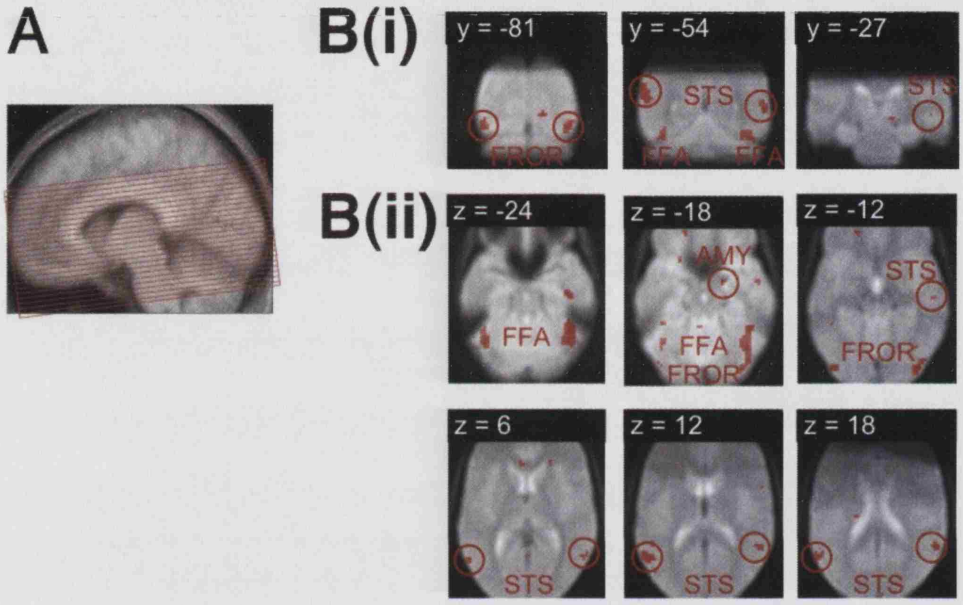

Figure 3.2: *fMRI coverage and results of localizer session*

a. Sagittal slice from group mean T1 structural image showing approximate location of 24 slices for functional imaging sequence. This sequence yielded coverage of inferior temporal cortex up to superior temporal sulcus, including amygdala, orbitofrontal cortex and fusiform cortices.

b. Coronal (i) and horizontal (ii) sections of group mean EPI showing mask derived from contrasts of faces>scrambled faces and emotional face>neutral faces. Note that the mask covers bilateral fusiform cortex (FFA), face-responsive occipital region (FROR), superior temporal sulcus (STS), and right amygdala (AMY).

interactions with emotion type were detected in regions of interest (see also Chapter 4); therefore the adaptation sessions were modelled without reference to different emotion types. Statistical threshold was set at $p<0.05$ corrected for multiple comparisons across a small volume of interest, using a mask derived from the localiser session (see below).

Mask defining regions of interest

Regions of interest were defined using two statistical results from the localiser session. First, effects of faces>scrambled faces were tested for and the result thresholded at $p<0.001$ uncorrected. Next, emotional faces>neutral faces was tested and this result thresholded at $p<0.001$ uncorrected. The two resulting masks were combined using an OR function, yielding the combined mask (Figure 3.2B) that was subsequently used for small volume correction (SVC, Worsley et al., 1996).

Visual similarity between face pairs

It might be argued that changes in identity and expression across pairs of faces do not represent categorical change alone, but also a variation in a continuous spectrum of visual similarity. Thus, for example, face pairs with the same identity and expression are likely to be more visually similar than pairs where identity is the same but the expression different. In order to avoid this potential confound affecting fMRI data analysis three measures of visual similarity were obtained and included in fMRI data analysis as covariates of no interest. The first two were image-based metrics, derived from mathematical analysis of image pairs comprising normalised least squares measures of differences between face pairs. Briefly, faces were normalised for

luminance and the second face subtracted from the first. The root mean square of the value at each pixel of this difference image was the difference score for a given image pair. In a refinement of the technique that accounted for minor differences in co-registration of salient features, faces were allowed to move over one another by up to 25 pixels in either plane and the minimum resulting value taken as the difference score (Vogels et al., 2001). The third measure adopted derived from an independent group of subjects (see below) who each saw 200 face pairs (50 of each trial type) presented with the same parameters as the imaging component of the study and rated each pair for visual similarity. The three ratings were in good agreement (Table 3.1). All three were included in the main statistical model as regressors of no interest by generating a design matrix in SPM whereby all events of interest were modelled as one trial type modulated parametrically by expansions to model the three similarity confounds. The columns pertaining to similarity confounds were then extracted and utilised in the model described above.

Analysis of eye-tracking data

fMRI differences in regions such fusiform cortex could be attributable to variations in visual attention with trial type (Wojciulik et al., 1998) and differences in emotionally-responsive regions such as amygdala might be similarly attributable to variations in arousal (Critchley et al., 2002). To explore such differences eye-tracking was used to monitor eye direction and pupil dilation. Data were acquired during scanning for a majority of subjects using an ASL504LRO eye-tracker (Applied Science Laboratories, Bedford, MA). Specifically, accurate pupillometry was achieved in nine subjects during the scanning session and accurate eye-gaze tracking in eight. Pupillometry data

	S_IS_E	S_ID_E	D_IS_E	D_ID_E
Computer measure 1	**0.86** (0.002)	**0.63** (0.013)	**0.44** (0.012)	**0.38** (0.019)
Computer measure 2	**0.84** (0.003)	**0.61** (0.011)	**0.44** (0.020)	**0.39** (0.014)
Human measure	**0.88** (0.005)	**0.53** (0.028)	**0.46** (0.014)	**0.24** (0.036)

Table 3.1: *Means (with standard deviations) of similarity measures for different event types*

Measures for pairs of stimuli were scaled from 0 to 1 and averaged across the trials used in the fMRI experiment. Higher values represent more similar face pairs. "Computer measure 1" represents the minimum value for the RMS difference between image pairs allowing a 25 pixel displacement in any direction; measure 2 represents the value with no displacement. The human measure results from a cohort of subjects who rated the similarity of face pairs on a visual analogue scale.

was analysed by defining a window of 1.2s after the second face and measuring the minimum, maximum and mean pupil diameter during averaged traces (low-pass filtered at 7.5Hz and baselined for the onset of the second face) from this window for each subject. These three measures were entered into separate 2X2 ANOVAs. Eye-gaze direction was also assessed using a summary statistic approach. For each of the four critical trial types spatial maps of eye-gaze density were constructed. Each of these maps was compared to the mean map, and difference images constructed. The root mean squares of the density difference values for these latter maps were entered into a 2X2 ANOVA.

Control data on explicit tasks for identity/expression detection

An important consideration is the possibility that subjects might not have noticed repetition of identity or expression of faces within each pair. Furthermore, if a change in one dimension (e.g. the expression of the faces) affected subjects' ability to detect repetition of the other dimension (e.g., the identity of the faces), then interactions between the two dimensions detected by fMRI would be difficult to interpret (in that a decrease in the fusiform response for S_ID_E trials relative to S_IS_E trials, for example, could simply reflect a reduction in the number of trials in which subjects were aware that a repeated face was of the same identity). Control behavioural experiments were conducted to test these possibilities.

An independent group of 16 subjects (age range=22-36, mean age=28.5; 11 males; 2 left-handers) completed three behavioural tasks, using identical procedural parameters to those in the imaging study. In the first task, they rated pairs of faces presented for

visual similarity using a computer-based visual analogue scale (providing the subjective measures of similarity mentioned above). In the second and third tasks, they classified face pairs as having the same or different identity, or the same or different emotional expression. The order of the identity/expression task and the response buttons used for same/different judgements were counter-balanced over subjects. In all, each subject performed 50 trials of each type for each task. A short (25 trial) practice session preceded each task.

Results

Behavioural data during scanning

Subjects detected (mean ± standard deviation) 99±2% of targets (chairs) in the localiser (false alarm rate = 0.2±0.4%), and 88±7% of targets (female faces) in the adaptation phase (false alarm rate = 7±6%).

The two measures derived from the eye-tracking data from the fMRI scanning sessions showed no significant differences between the four trial types of interest (for gaze direction all $F(1,7)<0.125$, all $p>0.7$; for pupil diameter all $F(1,8)<2.1$, all $p>0.18$, except a marginal trend for an interaction between identity and expression repetitions in the minimum pupil constriction: $F(1,8)= 4.0$, $p=0.08$). These non-significant results suggest that there were no detectable differences in visual attention (indexed by eye-gaze direction) or arousal (indexed by pupil diameter changes) between the different experimental conditions.

Neuroimaging data

Localiser phase

The two contrasts from the Localiser Phase (all faces > scrambled faces and emotional faces > neutral faces) were combined to create a mask of regions that responded to faces and/or facial expression. As expected, the activated regions included bilateral fusiform and more posterior occipital areas, as well as STS and amygdala and the mask is shown in Figure 3.2B. This mask defined a search region for the subsequent comparisons in the Adaptation Phase, allowing a principled means for correcting for multiple comparisons over voxels.

Adaptation phase: Main effect of repeated identity

As predicted, a significant main effect of repeated identity (reduced response when the second face exhibited the same identity as the first; $[D_I S_E + D_I D_E] > [S_I S_E + S_I D_E]$) was seen in right fusiform cortex (x,y,z = 39,-60,-15; Z = 3.76; $p<0.05$, one-tailed, small-volume corrected (SVC) for the localiser mask; Figure 3.3). In addition, adaptation for repeated identity was seen in right posterior STS (STSp) (x,y,z = 63,-51,15; Z = 3.73; $p<0.05$ SVC; Figure 3.4). Because an interaction or main effect of repeated expression would influence interpretation of these results, such effects were examined for at reduced threshold. In the peak right STSp voxel, a marginally significant main effect of repetition of expression was evident (Z=1.74; $p<0.05$, one-tailed, uncorrected), whereas in fusiform no such effect was evident (Z=0.99, $p>0.1$ one-tailed, uncorrected). There was no evidence for an interaction in the peak fusiform voxel (Z=1.26, $p>0.2$ two-

tailed, uncorrected), nor in the right STSp (Z=0.50, p>0.2 two-tailed). A more posterior region of right occipital cortex, possibly corresponding to a face-responsive occipital region (FROR), showed uncorrected repetition effects but failed to withstand correction for multiple comparisons across the volume of the mask (x,y,z = 42,-75,-18; Z = 3.59; p=0.071 SVC; see Figure 3.3).

Adaptation phase: Main effect of repeated expression

A region of right STS anterior to that described above was shown to be less active when the second face exhibited the same expression as the first face ($[S_1D_E+D_1D_E]>[S_1S_E+D_1S_E]$; Figure 3.5). This activation corrected for multiple comparisons across the volume of the mask (x,y,z = 57,-18,-12; Z = 3.80; p<0.05, one-tailed, SVC). Despite the apparent trend towards an interaction in Figure 3.5b, this was not significant when tested at a lenient statistical threshold (Z=1.06, p>0.2, two-tailed, uncorrected; Figure 3.5c).

Region-by-condition interaction

To determine whether the detectable differences in main effects between the mid-STS and fusiform regions were significant a region-by-condition interaction was tested using a 3mm sphere centred on each peak. A significant 3-way interaction obtained (p<0.05) in the direction predicted by the distributed model (adaptation to identity in fusiform and to expression in STS).

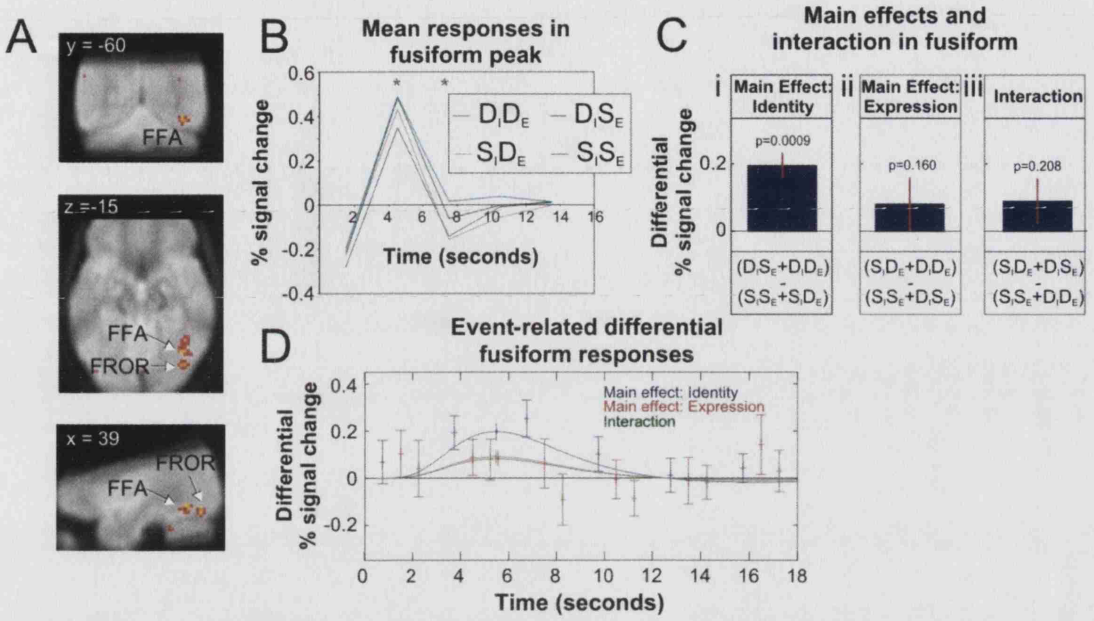

Figure 3.3: *Fusiform cortex shows fMRI-A for repeated identity*

a. SPM overlaid on group mean EPI showing activation in posterior occipital and fusiform cortices. Fusiform peak at x,y,z = 39,-60,-15; Z = 3.76. The more posterior activation (FROR) is at x,y,z = 42,-75,-18; Z = 3.59 but does not correct for the volume of the mask. Red = $p<0.01$ uncorrected, yellow = $p<0.001$ uncorrected. Results are displayed masked using the results from localizer scan (with the latter thresholded at $p<0.05$ uncorrected).

b. Mean response profiles for different event types from peak fusiform coordinates. Data derived from a finite impulse response (FIR) model with 3s time bins. Asterisks represent time points where the relevant main effect (adaptation for same identity trials) is significant at $p<0.05$ in the FIR model.

c. Differential effects in peak fusiform voxel using data from the main model for the three contrasts tested (main effect of identity (i), expression (ii) and their interaction (iii)). Bars represent the mean parameter estimate across subjects of the canonical haemodynamic response function; error bars represent standard error. p-values represent t-tests of the mean difference from zero. t-tests are one-tailed for main effects and two-tailed for the interaction term. Note that the interaction is **not** a comparison of the two main effects, and therefore would not be significant for an area selective for either identity or expression.

d. Differential responses in peak fusiform voxel with time. Differential effects derived from 3s time bin FIR model and from fitted responses of main model are shown (FIR datapoints for the main effect of identity are shifted backwards by 0.75s and for the interaction, forwards by 0.75s, for legibility).

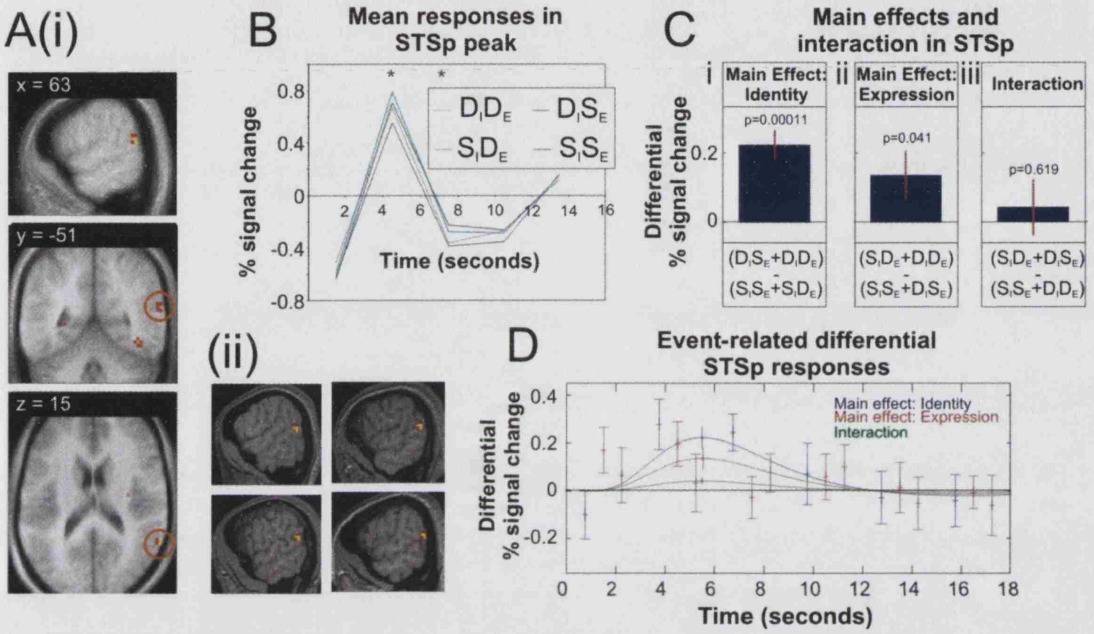

Figure 3.4: *Posterior STS shows fMRI-A for repeated identity*

a. Posterior STS (x,y,z = 63,-51,15; Z = 3.73) shows greater response to trials with different identities for the second face than repeated identities. This region appears to be around the posterior horizontal segment, though the anatomy is somewhat variable from subject to subject (ii). Display as in Figure 3.3a. (ii) shows group-result data shown on single subject T1 structurals to aid localisation.

b. Mean response profiles in peak posterior STS (STSp) voxel, derived from 3s time-bin FIR model. Display as in Figure 3.3b.

c. Differential effects in peak STSp voxel using data from the main model for the three contrasts tested. Display as in Figure 3.3c.

d. Differential responses in peak STSp voxel with time. Display as in Figure 3.3d.

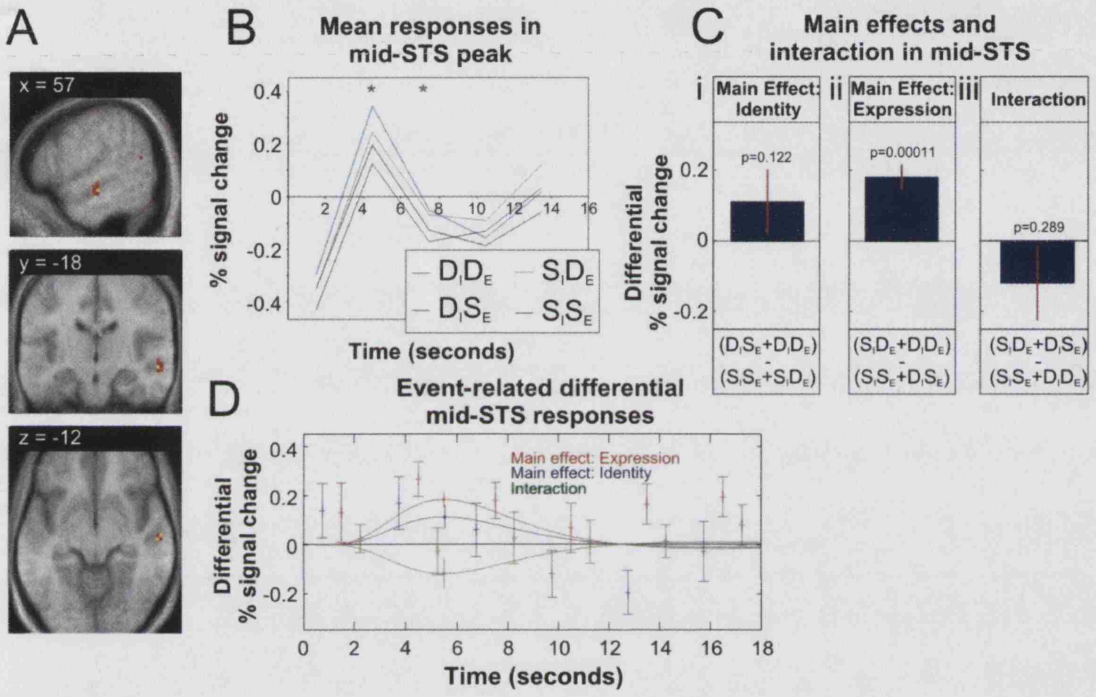

Figure 3.5: *Mid-STS shows fMRI-A for repeated emotion*

a. Mid-STS (x,y,z = 57,-18,-12; Z = 3.73) shows greater response to trials with different expressions for the second face than repeated expression. Display as in Figure 3.3a.

b. Mean response profiles in peak mid-STS voxel, derived from 3s time-bin FIR model. Display as in Figure 3.3b, except asterisks now represent significant differences for the main effect of emotion repetition.

c. Differential effects in peak mid-STS voxel using data from the main model for the three contrasts tested. Display as in Figure 3.3c.

d. Differential responses in peak mid-STS voxel with time. Display as in Figure 3.3d.

Adaptation phase: Interaction

No areas within the mask defining regions of interest showed an interaction between identity and expression.

Control Experiments

One potential confound in the experimental design is the presence of differences in the visual similarity between face pairs of different trial types. To account for this confound mean subjective and objective similarity measures for the four different trial types were obtained (Table 3.1). The subjective data were obtained from an additional behavioural experiment (see Methods). All three measures showed significant differences between the four trial types (all $F(1,13)>190$, all $p<0.001$). $S_I S_E$ pairs were more similar than the other three types, despite the use of different images in this condition. Unsurprisingly, and consistent with the concept of identity as an invariant feature of the face, trials with same identity had greater similarity ratings than trials with different identity. To account for these differences, all three measures were included as covariates of no interest in the analysis of the fMRI data (see Methods), which removed any linear contribution of similarity to the fMRI findings described above. A random effects analysis of the contribution of these regressors to the model (using an F-contrast spanning the three regressors in an ANOVA model) suggested that they were explaining effects in visual regions, though not within the mask used for SVC (e.g. peaks at: x,y,z = -51,-60,-6, Z = 3.76; x,y,z = 9,-78,9; Z = 3.46; x,y,z = 54,-45,-15, Z = 3.39; x,y,z = 30,-60,-15, Z = 3.38; all $p<0.001$ uncorrected).

An additional concern, noted above, is that subjects might not notice repetitions of identity or expression, or that the presence of repetition in one dimension would affect behaviour to the other dimension. Data from a behavioural experiments on a separate subject cohort (see Methods) showed that the mean accuracy in an identity discrimination task was 88% for trials when expression was held constant and 83% for trials when expression changed (paired t-test: $Z=3.78$, $p<0.001$). This was paralleled by slower reaction times (RTs) when judging identity in the context of expression changes (806ms vs 773ms, $Z=3.33$, $p<0.001$). Mean accuracy for the emotion discrimination task was 87% when identity was unchanged across the face pair and 80% when identity changed ($Z=3.53$, $p<0.001$). Again, RTs were slower on trials where the task-irrelevant dimension (identity) changed, compared to being held constant (888ms vs 848ms, $Z=4.41$, $p<0.001$). These data demonstrate that people's ability to detect repetition of identity or expression with these stimuli was generally high. The data also suggest that changes in one dimension do affect sensitivity to repetition of the other dimension. This behavioural interaction does not, however, confound the finding of two orthogonal main effects in the fusiform/STSp and mid-STS regions.

Discussion

In this experiment I used an event-related fMRI-adaptation paradigm to identify the neuroanatomical basis for coding different aspects of faces, specifically identity and expression. By presenting pairs of faces in which the identity and emotional expression of a second face were either concordant or varied with respect to the first, I demonstrated that discrete brain regions show a reduced BOLD signal when a specific dimension was repeated (relative to when it changed). Specifically, posterior lateral

right fusiform cortex and posterior right STS exhibited adaptation for identity, whereas right mid-STS showed adaptation for emotional expression. These differences do not relate to any obvious measure of visual similarity between faces in each pair, given that both objective and subjective measures of similarity were covaried out of fMRI data analysis. In addition, no evidence was found that the observed effects could be attributed to differences in eye-movement or arousal. Finally, control data showed that subjects' explicit ability to detect changes in identity or in expression was generally high. Although performance was reduced when the other dimension changed, which might confound any interaction between identity and expression on the levels of adaptation, this observation cannot explain the simultaneous finding of two orthogonal main effects in the imaging data.

In the distributed model of face processing (Haxby et al., 2000), there is a proposed dissociation between processing of invariant and changeable aspects of faces. Specifically, it is suggested that invariant features are coded in ventral occipital and lateral fusiform regions (also known as the "face area" - Kanwisher et al., 1997), whereas changeable aspects are coded by right STS. The fMRI data reported in this chapter broadly support this model. Within the framework of fMRI-A, the demonstration of a main effect of repeated identity in right fusiform cortex indicates this region represents identity, an invariant aspect of human faces. Although previous studies have shown repetition decreases to faces in fusiform cortex (George et al., 1999; Gauthier et al., 2000b; Henson et al., 2000, 2003), the present experiment is the first to show repetition effects in fusiform cortex across dramatically different views of the same identity (i.e., with different expressions). This finding is important because it suggests that face representations in this region encode not just a specific visual image

but a more abstract representation of facial identity. Two recent fMRI studies have reported similar decreases in BOLD signal with repetition of different images of the same identity (Vuilleumier et al., 2003a; Eger et al., 2004) but in these studies the different images were different spatial frequency bands from the same original stimulus. The current experiment and stimuli allow the broader conclusion that repetition of identity even across different views differentially engages fusiform.

A consistent finding in neuroimaging studies of emotional expression is activation of fusiform cortex during perception of emotional relative to neutral faces (Breiter et al., 1996; Dolan et al., 1996; Morris et al., 1998a; Vuilleumier et al., 2001, 2003a; Pessoa et al., 2002; Surguladze et al., 2003; see also Chapters 4 and 5). This finding has been interpreted as reflecting enhanced processing associated with arousing emotional faces relative to non-arousing neutral faces (Dolan, 2002). However, an alternative explanation is that this region encodes the emotionality of the face, resulting in enhanced activation when expressive faces are presented. The use of an adaptation paradigm in this study enables the potential dissociation between these possibilities. If this region coded for specific expressions, it should have shown adaptation for expression, akin to that for identity. The lack of evidence for adaptation for repeated expressions is consistent with the former interpretation that fusiform modulation is mediated by an amygdala-associated effect (though it should be noted that this inference is based on a null result). At the very least, it seems that any bottom-up effects of expression in right fusiform cortex are of less importance than those of identity, i.e. it exhibits relative preference for identity processing from faces.

In contrast to right fusiform, a focus in right mid-STS showed a main effect for repetition of emotional expression, with repeated expressions associated with reduced activation relative to differing expressions. This accords with a role for this region in coding the specific emotion expressed in a face. The anterior locus of this activation, which fell at –18 on the anterior-posterior axis, is somewhat surprising. Previous studies concerning facial expression have reported activation in right STS in a more posterior locus (around –35 to –60mm - Critchley et al., 2000a; Iidaka et al., 2001; Narumoto et al., 2001; see also Chapter 4). This more anterior locus is, however, within that sector of STS reported as activated in studies of social cues (Ojemann et al., 1992; Allison et al., 2000; Martin and Weisberg, 2003; Saxe and Kanwisher, 2003). In addition, activation was detected close to this locus in the study reported in Chapter 4 where a focus of activation was found in an explicit emotional judgement, relative to a gender judgement, task (for comparison: x,y,z = 52,-16,-18; Z = 3.96). Note also that this region fell within the face localiser mask and by this definition is responsive to faces or facial expression.

Posterior STS, like the fusiform, showed adaptation to repeated identity. This is contrary to a previous study that failed to observe repetition effects in this region (Henson et al., 2003), though that study used much longer repetition lags. Nevertheless, a role for posterior STS in processing personal identity is consistent with a recent human lesion study describing a patient with an infarct in the vicinity of left STS who described novel faces as familiar (Vuilleumier et al., 2003b). Unlike fusiform though, posterior STS showed a trend towards an additive main effect for repeated emotion, implying that its role in face processing may be multifaceted.

Intriguingly, in a recent re-analysis of single neuron data from monkeys, Tiberghien and colleagues (2003) suggest that all facial features contribute to distinguishing identity, whereas only a subset determine facial expression. These authors hypothesise that as a consequence, inferior temporal regions in monkeys may contain identity-selective neuronal populations, whilst STS might contain populations sensitive to identity and expression. Such a view fits with the demonstration in this experiment of sensitivity to identity-repetition in posterior STS and sensitivity to expression in posterior and mid-STS. However, with regard to the human lesion literature, the majority of reported prosopagnosic patients described have inferior occipitotemporal rather than lateral temporal lesions (see e.g. Damasio et al., 1990a; Wada and Yamamoto, 2001), presumably corresponding to fusiform rather than STS (but see Figure 1A in Tranel et al., 1997; see also Rossion et al., 2003). In addition, monkeys with STS lesions appear to have only minor identity discrimination deficits (Heywood and Cowey, 1992), and recent evidence from multidimensional scaling analysis of single neuron data from monkey STS and inferior temporal (IT) cortex also suggests that STS is more concerned with analysis of facial view while the code in IT is more concerned by facial identity (Eifuku et al., 2004). These apparent discrepancies with posterior STS responsivity to identity might be explained in a number of ways. Firstly, it is possible that activation in this region is epiphenomenal and of no functional consequence for identity recognition. However, another possibility is that the stimuli used in the current experiment tax identity processing across different views of a face, given that the role of STS in processing different views of face stimuli is well known (Perrett et al., 1985, 1991; Eifuku et al., 2004). In addition, it has been demonstrated that STS neurons in the macaque monkey process identity, at least in the form of a population code (Baylis et al., 1985), and there is evidence that some single neurons in STS code for the same

identity across different face views while other STS cells code conjunctions of identity and view (Perrett et al., 1991). It may be the case that the aspect(s) of identity processing that occurs in STS are insufficient when compromised to engender a full blown prosopagnosia. It follows also that tests designed to probe prosopagnosia may be relatively insensitive to these aspect(s) of identity processing expressed in STS. In addition, it is striking to note that cortical stimulation of a lateral site (in posterior middle temporal gyrus, adjacent to STS) of N200 face-specific responses resulted in temporary impairment in face recognition (Puce et al., 1999 p 451).

Previous work has implicated other brain regions in processing facial expressions, most notably the amygdala (Breiter et al., 1996; Morris et al., 1996; Whalen et al., 1998; Vuilleumier et al., 2001, 2003a; Pessoa et al., 2002; see also Chapter 4). There are a number of potential reasons why adaptation was not detected in this region. One possibility is that amygdala responses are emotion-specific, with greatest responses to fearful faces (Calder et al., 2001), and thus collapsing across different expression sub-types may have obscured responses in this region. Alternatively, the amygdala might code for facial expression in a different manner from cortical regions such as STS, with a non-specific code whereby responses are dependent upon the arousal engendered by the emotion (Critchley et al., 2005). Another possibility is that an expression-specific amygdala response may be insensitive to adaptation, though this seems unlikely, given positive findings concerning the amygdala and stimulus repetition (Ono and Nishijo, 2000; Rotshtein et al., 2001).

Although identity and emotion may be processed by partially dissociable neural pathways, the two pathways are likely to interact in production of behavioural

responses. This would appear to be the case for the explicit identity and emotion detection tasks described, in which a change in one dimension (identity or emotion) impaired ability to detect changes in the other dimension. In behavioural studies, other studies have also found evidence for the non-independence of identity and emotion processing (Schweinberger and Soukup, 1998; Schweinberger et al., 1999; Ganel and Goshen-Gottstein, 2004). A further brain region may be responsible for the integration of these distinct aspects of face processing which was not detected in this study. A more explicit behavioural task during fMRI may help to clarify this issue in future studies.

One issue that deserves consideration is the meaning of BOLD changes in adaptation paradigms such as this. It has been demonstrated that local field potentials (LFP) correlate with the BOLD signal better than multi- or single- unit activity in the macaque monkey (Logothetis et al., 2001). Thus, a region showing fMRI-adaptation may not be transmitting fewer spikes but may either be showing a reduced afferent input or reduced local processing. This highlights one possible dissociation between fMRI-adaptation and response suppression as recorded in single unit work in monkeys (see Desimone, 1996). However, fMRI experiments based upon adaptation are not uniquely problematic in this regard; instead, this a more general interpretational issue for unifying electrophysiological and fMRI work. (See Henson and Rugg (2003) for a more extensive discussion of haemodynamic decreases and response suppression.)

Summary

The experiment reported in this chapter provides the first direct evidence in healthy human subjects for a neuroanatomical dissociation between identity and expression processing. Specifically, I show that fusiform cortex demonstrates fMRI adaptation when the identity of a face is repeated, and a region of STS shows adaptation when the emotional expression of a face is repeated. The response profiles of these two regions were significantly different in the directions predicted by the distributed model of face processing (Haxby et al., 2000), and the findings are generally consistent with this model. However, an adaptation response in posterior STS to repeated identity suggests that STS may also manifest a degree of functional segregation in face perception. Differential adaptation responses in fusiform and STS demonstrate that emotive aspects of face perception are dissociable from regions involved in identity processing.

Chapter 4: Neural systems responsive to facial emotion

Introduction

The experiment described in the previous chapter demonstrated that regions encoding facial identity are dissociable from those encoding facial expressions of emotion. In this chapter, I describe an experiment designed to explore further the regions encoding facial emotion. Specifically, the issues of dissociable brain regions encoding different emotions and the task-dependence of brain regions responsive to facial emotion are investigated.

Social cognition reflects an ability to recognise and respond appropriately to emotion in others (Brothers, 1990; Adolphs, 1999). A pan-cultural means of signalling emotion is through facial expression (Darwin, 1872; Ekman, 1993) and the successful interpretation of facial expressions of emotion has been considered to constitute an innate, non-verbal, automatic process (Darwin, 1872; Ohman, 2002). An influential contemporary theory of human emotion, the "somatic marker hypothesis", proposes an intimate association between bodily and brain states in the perception and experience of emotion (Damasio, 1994, 1999).

The somatic marker hypothesis has been extended in a recent proposal that perception of an emotion in a conspecific involves simulation of the emotional state within the relevant cortical circuitry of the observer (Adolphs et al., 2000; Adolphs, 2002a, b). Support for this draws largely upon evidence from studies of patients with discrete brain

lesions. In one of these studies, patients with ventral prefrontal lesions showed deficits in recognition of emotion from faces and voices relative to patients with lesions at other sites (Hornak et al., 1996). Ventral prefrontal cortex has long been implicated in the generation and mapping of somatic markers (summarised in Damasio, 1994, 1999). A study of a larger sample of patients suggested that right somatosensory cortices were particularly critical to this function (Adolphs et al., 2000). One problem in relation to these findings is that neuroimaging studies have generally failed to demonstrate activation in these regions during facial emotion perception. A possible reason is that human neuropsychological studies might conflate inability to perform the recognition task with impaired perception of facial emotion.

A further issue in the study of emotion perception is whether different brain regions process distinct emotions. Recent neuropsychological and neuroimaging data have been interpreted as indicating that emotional perception, and perception of facial expressions in particular, is organised in a modular fashion with distinct neural circuitry subserving individual emotions (Calder et al., 2001; Adolphs, 2002b). However, examination of these data suggests that this conclusion may be premature, and whether there is a common substrate to the perception of multiple basic emotions has not been systematically investigated using non-invasive neuroimaging techniques.

Much early research into brain mechanisms for emotional perception involved either single unit recording in non-human primates or neuropsychological studies of patients with focal lesions. Neurophysiological evidence suggests that circumscribed brain regions contain neurons responsive to facial movements and specific facial expressions, and that these neurons are segregated from those responding to facial identity

(Hasselmo et al., 1989; see also Chapter 3). Lesion studies in humans report two distinct findings: certain lesions impair recognition of all so-called "basic emotions" (Adolphs et al., 1996, 2000; Hornak et al., 1996) whereas lesions in other brain areas impair recognition of a more limited subset of facial emotions (Adolphs et al., 1994, 1995; Calder et al., 1996; Sprengelmeyer et al., 1999). Thus, ventral prefrontal and right cortical lesions result in a generic impairment in facial emotion recognition as measured in group studies (Kolb et al., 1983; Rapcsak et al., 1993; Adolphs et al., 1996, 2000; Hornak et al., 1996). As indicated above, right somatosensory cortex has been highlighted as a critical locus in facial emotion recognition that spans different emotions in the largest group study to date (Adolphs et al., 2000). Single subject data, by contrast, has focused on consequences of damage to distinct brain regions causing impairment in recognition of specific emotions. For example, impairment in fear recognition is reported following bilateral damage to the amygdala (e.g. patients NM, RB, SE, SM and YW - Adolphs et al., 1994, 1995; Calder et al., 1996; Broks et al., 1998; Sprengelmeyer et al., 1999), but the selectivity of this deficit remains unclear insofar as patients are commonly impaired on recognition of more than one emotion (e.g. patients DR, GP, JC, JM, NM, and SP – Young et al., 1995, 1996; Calder et al., 1996; Broks et al., 1998; Adolphs et al., 1999; Rapcsak et al., 2000; Schmolck and Squire, 2001), and occasionally do not manifest fear recognition deficits (e.g. patients DBB, EP, GT and RH - Hamann et al., 1996; Adolphs et al., 1999). Additionally, a single patient with discrete lesions in left insula and basal ganglia has been described as showing a specific deficit in disgust recognition and experience (Calder et al., 2000). Although a double dissociation between neural substrates of recognition for fear and disgust is of considerable theoretical interest (Calder et al., 2001), caution has been

urged on the basis that individual differences between subjects and task difficulty may confound inferences from single case studies (Rapcsak et al., 2000; Adolphs, 2002a).

A powerful means of studying the neural correlates of emotion perception in the intact human brain is afforded by functional neuroimaging. To date, studies indicate that discrete brain regions process human facial emotions even when facial emotion perception is not task relevant, concomitant with the idea that facial expression is processed automatically (Breiter et al., 1996; Dolan et al., 1996; Morris et al., 1996; Phillips et al., 1997, 1998; Whalen et al., 1998; Blair et al., 1999; Vuilleumier et al., 2001). Such findings raise the possibility that regions involved in perception of facial emotion may be dissociable from those involved in acting upon that information. This is an important consideration insofar as patient lesion studies reflect impairments on direct emotion recognition tasks. The most widely replicated neuroimaging finding is an amygdala response to fearful faces (Breiter et al., 1996; Morris et al., 1996; Phillips et al., 1997, 1998; Whalen et al., 1998; Hariri et al., 2000; Dolan et al., 2001; Vuilleumier et al., 2001; Pessoa et al., 2002). Facial expressions of disgust are reported to activate insula and basal ganglia, but not amygdala (Phillips et al., 1997, 1998). However, the claim that neuroimaging and lesion data have established dissociable neural substrates for fear and disgust (Calder et al., 2001) is open to question. For example, fearful faces are reported to activate insula and basal ganglia (Phillips et al., 1999) while disgust faces are reported to activate amygdala more than happy (Gorno-Tempini et al., 2001) and neutral faces (Phillips et al., 1998). In addition, comparing patterns of activation for disgust relative to neutral faces and for fearful relative to neutral faces with data derived from different studies does not provide a statistically convincing means of evaluating differences between emotions. Inferences regarding

dissociable responses to distinct emotions require a direct comparison of different emotions within the same study, and few such studies have been performed. In one such study with six subjects, Phillips and colleagues (1998) reported that certain regions showed greater responses to disgust than to fear within left anterior insula and right putamen/globus pallidus. In contrast, these authors also reported greater activation to fear than disgust in amygdala, though the coordinate for this activation corresponds better to anterior hippocampus than amygdala. Kesler/West and colleagues (2001) also examined responses to multiple emotions within one experimental paradigm and, using a region of interest approach, demonstrated dissociable responses to different emotional expressions (angry, fearful, happy and sad) with a greater response to fearful faces in inferior frontal gyrus and to angry faces in medial frontal cortex. Limitations of this study include the absence of a constrained experimental task and the use of a block design, with different emotions presented in different runs, rendering it impossible to separate effects of emotional expression from cognitive set and anticipatory effects engendered by block presentation.

The approach adopted in this experiment is to assess the effect of task (direct versus incidental processing) on neural responses to facial expressions when multiple emotions are considered within a single experiment. As highlighted above, the issue of task effects is critical in reconciling human lesion with functional imaging data. Lesion studies are limited to tasks requiring explicit recognition of facial emotion, whereas the majority of functional imaging studies have relied upon incidental emotional processing. A limitation of lesion studies is the possibility that they may conflate subjects' inability to utilise intact perception of facial emotion to successfully perform a task with an inability to accurately perceive the emotion. Two imaging experiments

have tackled the question of task effects (Critchley et al., 2000a; Gorno-Tempini et al., 2001) though both were limited by blocked presentation of emotional stimuli. This entailed subjects making judgements of emotion ("emotional or non-emotional") within a block of either emotional or non-emotional faces, a sub-optimal design. Here, an optimal approach offered by event-related designs is adopted, wherein stimulus order is randomised such that subjects are unable to predict the next stimulus from the preceding one. In order to assess the neural correlates of making judgements concerning the emotional expression of faces the design incorporated an "incidental" task in which subjects made a complex gender judgement concerning a pair of faces and a task in which subjects judged emotion in a pair of faces ("direct" task).

Methods

Stimuli

Stimuli were generated from facial photographs of eight individuals from the Ekman and Friesen (1975) series of emotional faces. Commercially available morphing software (Morpher 2 provided by M. Fujimiya; http://www.asahi-net.or.jp/~FX6M-FJMY/mop00e.html) was used to generate stimuli. Neutral faces of one gender were morphed towards emotionally expressive faces of the other gender. Stimuli were generated at 10%, 30%, 70% and 90% between the two original images. The morphing procedure involved delineating anatomical loci and regions common to the two original images. In total, approximately 140 loci were delineated per face, to include the outline of the face, forehead wrinkles, the eyes and eyebrows, orbicularis occuli, the nose, zygomaticus muscles, and chin (Figure 4.1a). Four emotions were used as target

morphs: disgusted, fearful, happy and sad faces. The identities used to generate the faces were: C, MO, NE, SW (females) and EM, JJ, PE, WF (males). Example stimuli are shown in Figure 4.1b.

Subjects

Twelve right-handed volunteer subjects consented to take part in the study, which was approved by the Joint National Hospital for Neurology and Neurosurgery/Institute of Neurology Ethics Committee. Subjects were free from medication, neurological or psychiatric history, and ranged in age from 22 to 41. Data from one scanning session in one subject was lost due to technical failure, and as a consequence this subject's data was not included in the group analysis. The group analysis therefore consisted of data from eleven subjects (five females) with mean age of 26.

Task

Subjects saw two faces either side of a fixation cross which remained on screen during the intertrial interval. Faces were either 10% and 30% or 70% and 90% morphs from the same series of images (Figure 4.1b). Trials containing 10%/30% morphs are referred to as 'low intensity' trials and trials consisting of 70%/90% morphs are referred to as 'high intensity' trials. Subjects' task during the experiment was blocked and alternated between choosing which face was more emotional or which face was more masculine (Figure 4.1c,d). Faces appeared for 1.5s and minimum stimulus onset asynchrony was 3s. Blocks consisted of ten trials of one task followed by ten trials of

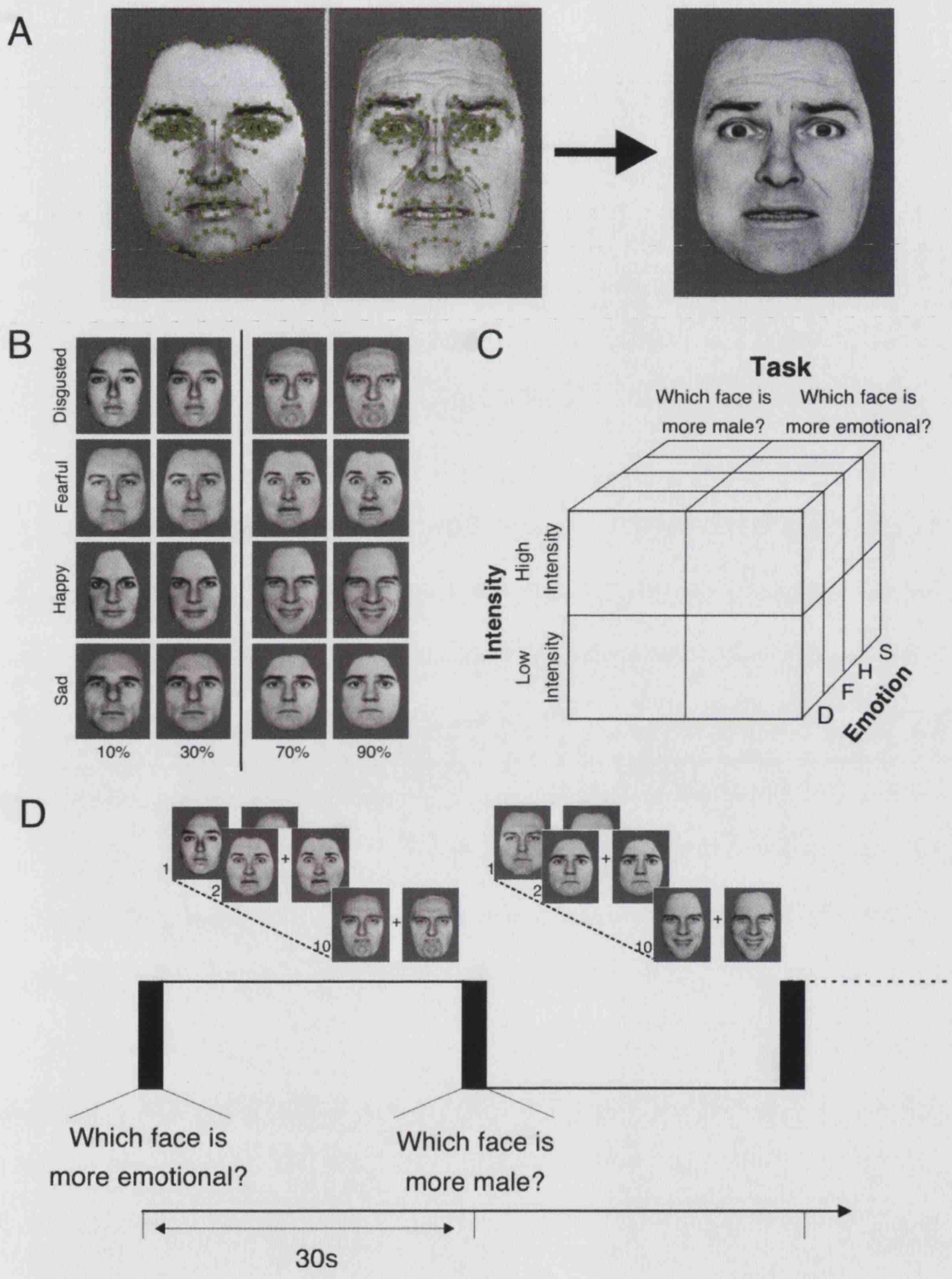

Figure 4.1: *Study design and example stimuli*

(Legend on following page)

Figure 4.1: *Study design and example stimuli*

a. Schematic of morphing procedure used to generate stimuli. Corresponding anatomical loci were delineated on two original face stimuli of opposite gender. The morphing software warps corresponding points and cross-fades textures to generate life-like intermediate stimuli that were used in the experiment.

b. Examples of stimuli used in the experiment. Stimuli were morphs between four male and four female identities from the Ekman and Friesen (1975) series. Morphs were generated between neutral faces of one gender and emotionally expressive faces (either disgusted, fearful, happy or sad) of the other at 10, 30, 70 and 90% along the spectrum between the two starting images.

c. Experimental design: Trials consisted of pairs of faces from the same morph spectrum wherein faces were either 10 and 30% emotionally expressive ("Low intensity" trials) or 70 and 90% emotionally expressive ("High intensity" trials). The task was either a judgement of which face was more masculine ("incidental" trials) or which face was more emotional ("direct" trials). Faces expressing four different emotions (D=disgust, F=fear, H=happy, S=sad) were used, engendering a 4X2X2 factorial design.

d. Example series of stimuli and blocking of task. Stimuli were presented as pairs of faces separated by a fixation cross. Task was blocked, each block being preceded by the appropriate task instruction. Blocks consisted of ten events (stimulus pair or null events). Minimum SOA was 3s; blocks were 30s long.

the other. Changes of block were indicated to subjects by means of a text display at the start of each block indicating which decision they were to make. Within each block, low and high intensity trial types of each emotion were randomly intermixed with one another and with null events. Scanning was divided into two 13-minute sessions. Within each session 192 events (12 events of each type) were randomly interspersed with 48 null events. The total number of events of each type was therefore 24. The first task block was counterbalanced across subjects.

Prior to scanning, subjects were shown examples of stimuli similar to those used during the scanning session and were trained on the experimental task.

fMRI scanning

Subjects were scanned during task performance using a Siemens VISION system at 2 Tesla to acquire gradient-echo, echoplanar T2*-weighted images with BOLD (blood oxygenation level dependent) contrast. Each volume comprised 33 X 3.3mm axial scans with 3-mm in-plane resolution. For each of two sessions, 340 volumes were continuously acquired every 2.5s. Subjects were placed in light head restraint within the scanner to limit head movement during acquisition. To allow for T1 equilibration effects the first five volumes of each run were subsequently discarded. Additionally a T1-weighted structural image was acquired in each subject. These structural images were coregistered with the mean EPI from the functional acquisition, normalised into a standard space using the normalisation parameters applied to the EPIs and subsequently averaged for overlay of statistical parametric maps.

During scanning, eye position was monitored using an infra-red eye tracker (ASL Model 450, Applied Science Group Co., Bedford, MA; refresh rate=60Hz). This data was available for eight subjects. Due to technical failure, temporal resolution was inadequate for confident trial-by-trial analysis of eye position. However, it was possible to compare eye position across the block component of the task. This was undertaken by defining a 25 second time window and measuring the cumulative position along the horizontal axis in three locations approximating to the left face, right face and fixation cross. Differences (between incidental and direct task) in cumulative time spent at each locus during were then entered into a series of one-sample T-tests.

Data analysis

Preprocessing and statistical analysis of imaging data was undertaken using SPM99 using the standard approach described in the general methods chapter. In statistical analysis, regressors corresponding to linear trends in response with time were constructed for each event type given the widely described phenomenon of changes in the profile of responses to emotionally-salient stimuli with time (e.g. Breiter et al., 1996; Buchel et al., 1998b, 1999; Phillips et al., 2001), though data from these effects are not reported here. For random effects analysis, one sample t-tests or ANOVAs were adopted, as appropriate. Where ANOVAs were used, departures from sphericity assumptions were accommodated.

A global threshold was set at $p<0.001$ uncorrected for multiple comparisons. Results that survive corrections for multiple comparisons across the whole brain at $p<0.05$ are reported, and $p<0.001$ uncorrected in regions predicted *a priori*. Additionally, it is

indicated whether activations in predicted regions survived a correction for multiple comparisons across small volumes of interest (Worsley et al., 1996). In right posterior STS, small volume correction (SVC) was based upon a sphere centred upon a previously established coordinate (12mm radius from the coordinate in Chapter 6: x,y,z=56,-44,4). In the cases of right SI and bilateral amygdala and insula masks (volumes $20cm^3$, $8cm^3$, $40cm^3$ respectively) based upon on the group's average structural scan were constructed using MRIcro (Rorden and Brett, 2000). For fusiform and face-responsive extrastriate cortex a mask (volume $35cm^3$) derived from an independent study was used (faces versus fixation), extending from –84mm posteriorly to -40mm anteriorly, based on the maximum and minimum coordinates typically reported for face-specific activations (Kanwisher et al., 1997; Haxby et al., 1999; Vuilleumier et al., 2001).

Results

Behavioural

Participants' reaction times were analysed by means of repeated-measures ANOVA using SPSS (SPSS Inc., Chicago). This analysis revealed no main effect of task or intensity ($p>0.7$), but a main effect of emotion was significant ($F_{(3,30)}=2.95$, $p<0.05$), clarified by a significant emotion-by-task interaction ($F_{(2.0,19.7)}=10.05$, $p<0.05$). This latter effect seemed driven by a facilitation effect in the direct task specific to happy faces and a cost in the incidental task specific to fearful faces (Figure 4.2). No higher order interaction with intensity was evident ($p>0.2$).

Cumulative time spent looking at the left face, right face and fixation cross were measured and compared across tasks for eight subjects for whom eye-tracking data were available. No significant difference between tasks was found for any of these three measures (left face: mean difference=-75.8; t_7=-1.47; p=0.18; right face: mean difference=123.7; t_7=1.40; p=0.20; fixation cross: mean difference=-152; t_7=-1.78; p=0.11).

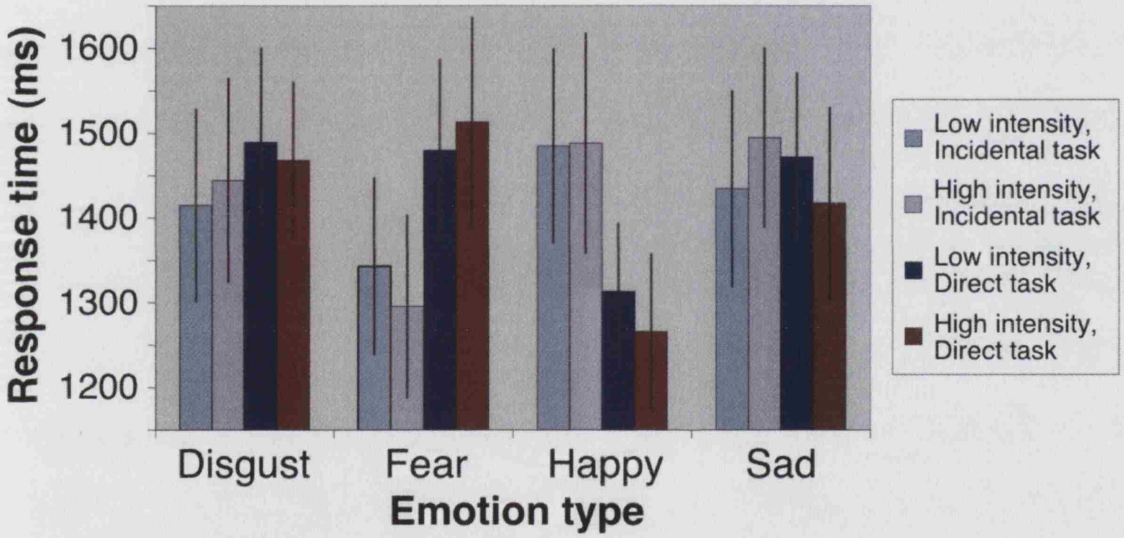

Figure 4.2: *Reaction time data across conditions*

Mean reaction time for each condition, error bars represent standard error. There is a significant main effect of emotion type, clarified by the interaction between emotion type and task. This effect appears to result from faster responses in the direct task for happy faces and in the incidental task for fearful faces.

Neuroimaging

Main effect of intensity

The contrast of high versus low intensity expressions (collapsing across task and emotion type) showed activation in predicted regions: fusiform and extrastriate cortex (x,y,z=34,-72,-20, Z=3.99, $p<0.1$ SVC; x,y,z=-36,-46,-24, Z=3.73, $p<0.001$ uncorrected) and bilateral amygdala (x,y,z=34,0,-26, Z=3.50, $p<0.05$ SVC; x,y,z=-24,-6,-18, Z=3.28, $p<0.001$ uncorrected). To ensure these activations were not a result of a disproportionate contribution from one task, conjunction analysis (Price and Friston, 1997) was performed, using the contrasts of high minus low intensity expressions across the two tasks (in effect testing the task independence of these responses). Significant bilateral amygdala and fusiform responses were evident in this contrast confirming the task-independence of these activations.

Tests for similarities between emotions – conjunction analysis

Although amygdala and face-responsive regions of extrastriate and inferior temporal cortex were activated by the main effect of intensity (see above), such analysis does not preclude the possibility of a predominant effect from one or more distinct emotions. A conjunction analysis across the contrasts of high intensity minus low intensity trials (collapsed across task) from each individual emotion type was thus carried out. In effect this conjunction analysis examines for a common effect across four contrasts, i.e. the effect of highly emotional faces, independent of sub-type of emotion. This analysis revealed significant activations in bilateral amygdala (x,y,z=30,-4,-22, Z=3.77, $p<0.05$ SVC; x,y,z=-24,-2,-24, Z=3.24, $p<0.001$ uncorrected; Figure 4.3) as well as fusiform

and extrastriate regions (x,y,z=28,-66,-12, Z=4.57, p<0.05 SVC; x,y,z=-30,-42,-20, Z=4.26, p<0.05 SVC; Figure 4.4). On the basis of this finding, the null hypothesis that at least one of these emotions failed to activate the amygdala can be rejected. A trend towards significance was evident in right posterior STS (x,y,z=48,-46,4; Z=3.52; p<0.1 SVC; Figure 4.5).

It should be noted that this analysis collapses across tasks and does not therefore preclude the possibility that one task does not contribute to this effect of intensity. A further conjunction analysis across the eight contrasts representing high intensity minus low intensity for each emotion category under each task condition was thus carried out. Even under these stringent conditions activation in left amygdala and bilateral fusiform and extrastriate regions (x,y,z=-26,-2,-22; Z=3.31; x,y,z=-38,-46,-18; Z=3.81; x,y,z=46,-60,-22; Z=3.18; all p<0.001 uncorrected) remained, indicating the effect of high intensity emotion was emotion type- and task-independent.

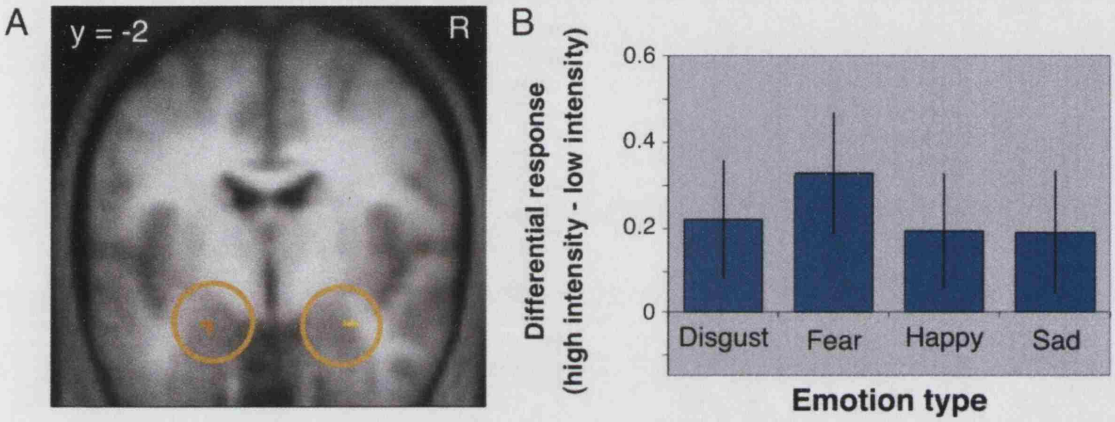

Figure 4.3: *Amygdala response common to distinct facial emotions*

a. Bilateral amygdala activation in the conjunction of high intensity-low intensity across emotions (collapsed across task). $SPM_{[t]}$ thresholded at $p<0.001$ uncorrected overlaid on averaged T1 scan of participants. Right peak at: $x,y,z=30,-4,-22$; $Z=3.77$; Left peak at: $x,y,z=-24,-2,-24$; $Z=3.24$.

b. Consistency of effect in right amygdala across emotions. y-axis represents size of differential response to high and low intensity expressions in amygdala (% signal change). Note that although there is numerically a greater response to fear than to the other emotions, this trend is absent in other voxels within amygdala and does not attain statistical significance ($p>0.8$). Importantly, if fear is omitted from the conjunction, amygdala voxels still obtain from this analysis.

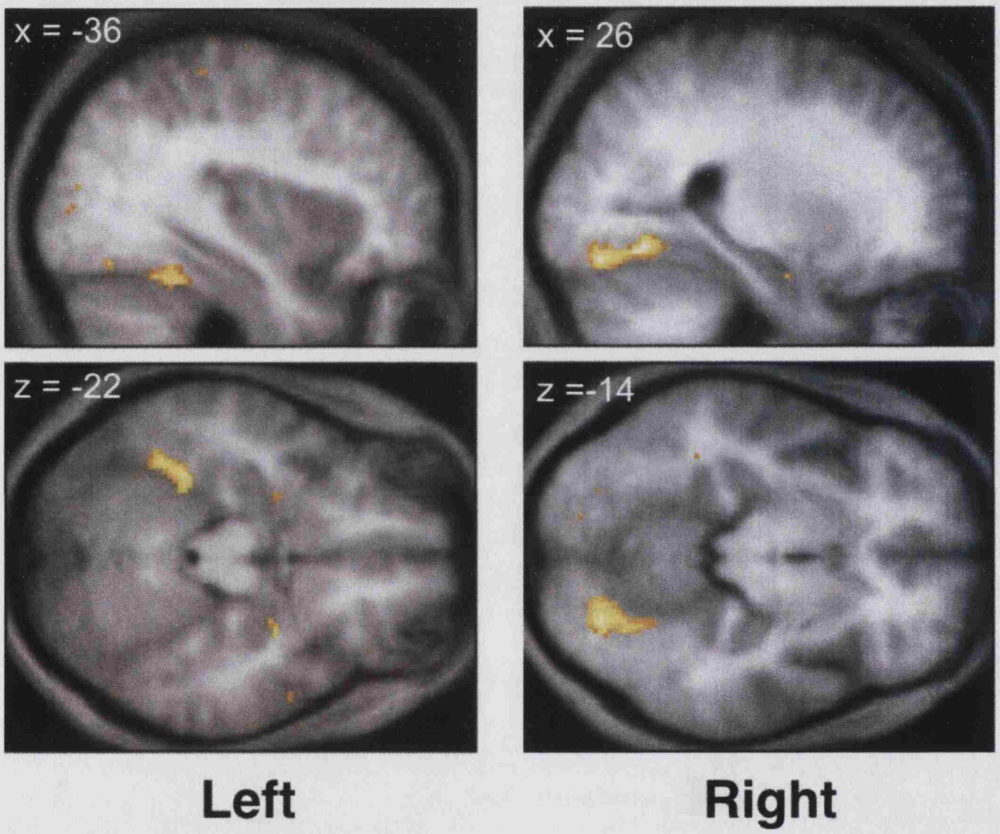

Figure 4.4: *Fusiform and extrastriate activations common to distinct facial emotions*

Bilateral fusiform and extrastriate cortex activation in the conjunction of high intensity minus low intensity emotion expression across emotions (collapsed across task). Right peak at: x,y,z=28,-66,-12; Z=4.57; Left peak at: x,y,z=-30,-42,-20; Z=4.26 Display as in Figure 4.3a.

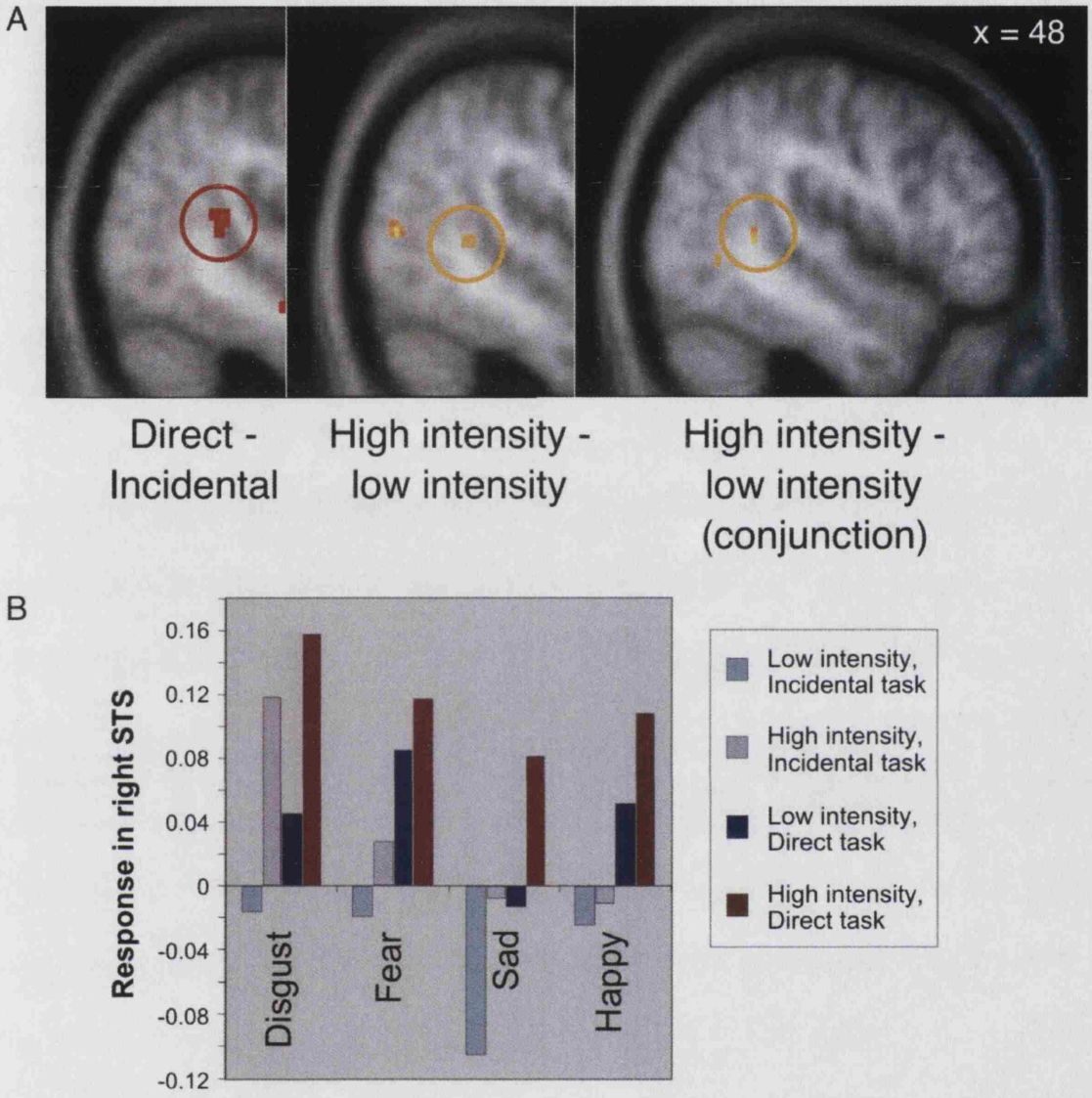

Figure 4.5: *Responses in right posterior STS*

a. Response in right posterior STS to task (direct>incidental), intensity (high>low) and conjunction of simple effects of intensity (high>low) across emotions. Display as in Figure 4.3a.

b. Response to individual event types in right STS. Note the increased response to high intensity expressions for each emotion and the additive effect of task, wherein responses are greater in the direct task. y-axis represents size of effect in peak voxel (% signal change).

Main effect of task

In the contrast of judging emotion versus gender (i.e. the comparison of direct versus incidental emotional processing, collapsed across emotion sub-type and intensity), activation was observed in medial prefrontal cortex/anterior cingulate (x,y,z=-16,42,-8, Z=5.45; Figure 4.6) and posterior cingulate/retrosplenial cortex (x,y,z=-8,-58,16, Z=5.34; Figure 4.7) [both $p<0.05$ corrected for multiple comparisons]. Importantly, however, this latter region was also activated in the interaction between task and intensity (see below).

In view of the strong *a priori* hypothesis of involvement of somatosensory cortices in explicit emotion judgement, a striking finding was activation in right somatosensory cortex (x,y,z=44,-12,48, Z=4.02, $p<0.05$ SVC; Figure 4.8). Additionally, at uncorrected thresholds of $p<0.001$, activations in predicted regions were evident in bilateral insula (x,y,z=40,-24,4, Z=4.32; x,y,z=-40,-16,-8, Z=3.38) and SII (x,y,z=-56,-12,12, Z=3.85) though the former failed to correct for multiple comparisons across bilateral insula volume. Finally, activation was seen in STS (posterior focus: x,y,z=58,-34,8, Z=3.91, $p<0.05$ SVC; anterior focus: x,y,z=52,-16,-18; Z=3.96; Figure 4.5).

Conjunction analyses across the distinct emotions for the task effects confirmed that regions highlighted in the main effect of task were activated independent of emotion sub-type. In the reverse contrast (judging gender versus judging emotion) there were no significant effects at corrected thresholds.

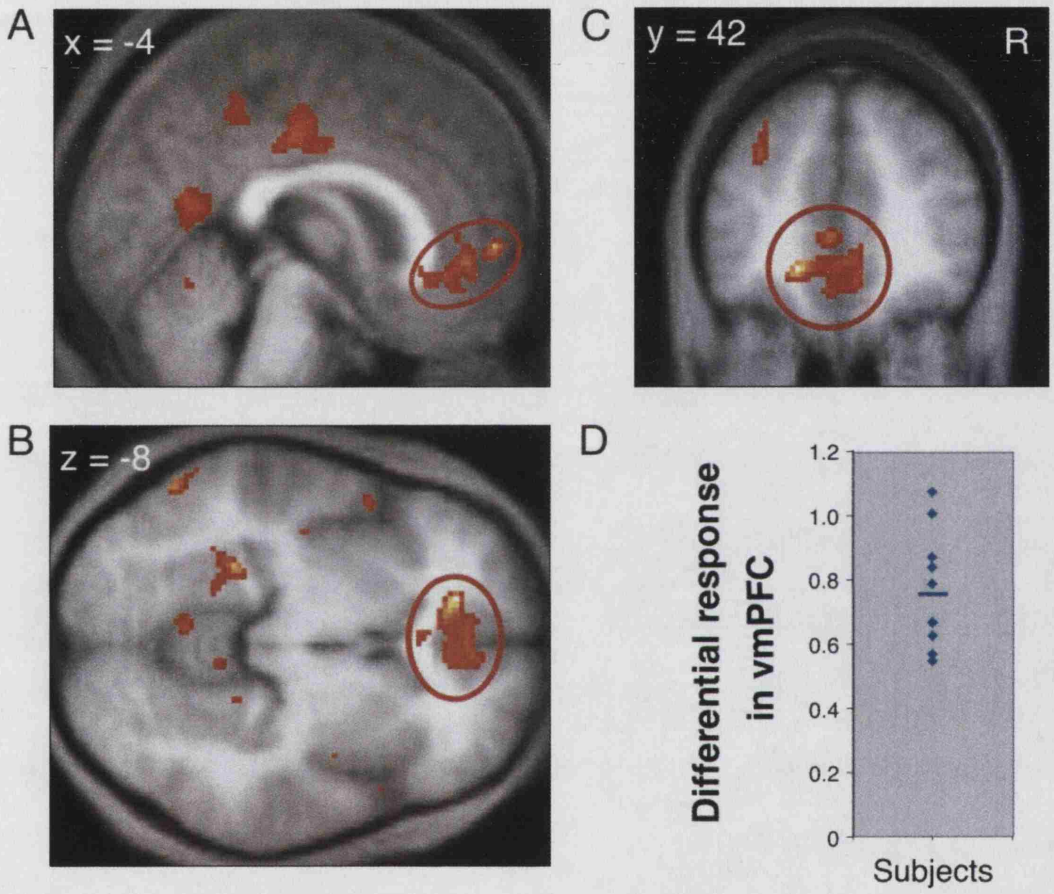

Figure 4.6: *Ventromedial prefrontal cortex responses in direct task*

a-c. Sagittal, horizontal and coronal sections showing highly significant response during direct task compared to incidental task in ventromedial prefrontal cortex (vmPFC). Display as in Figure 4.3a.

d. Consistency of response in peak voxel in vmPFC. Diamonds represent individual subjects, horizontal line, the mean response. y-axis represents size of differential response in vmPFC (% signal change).

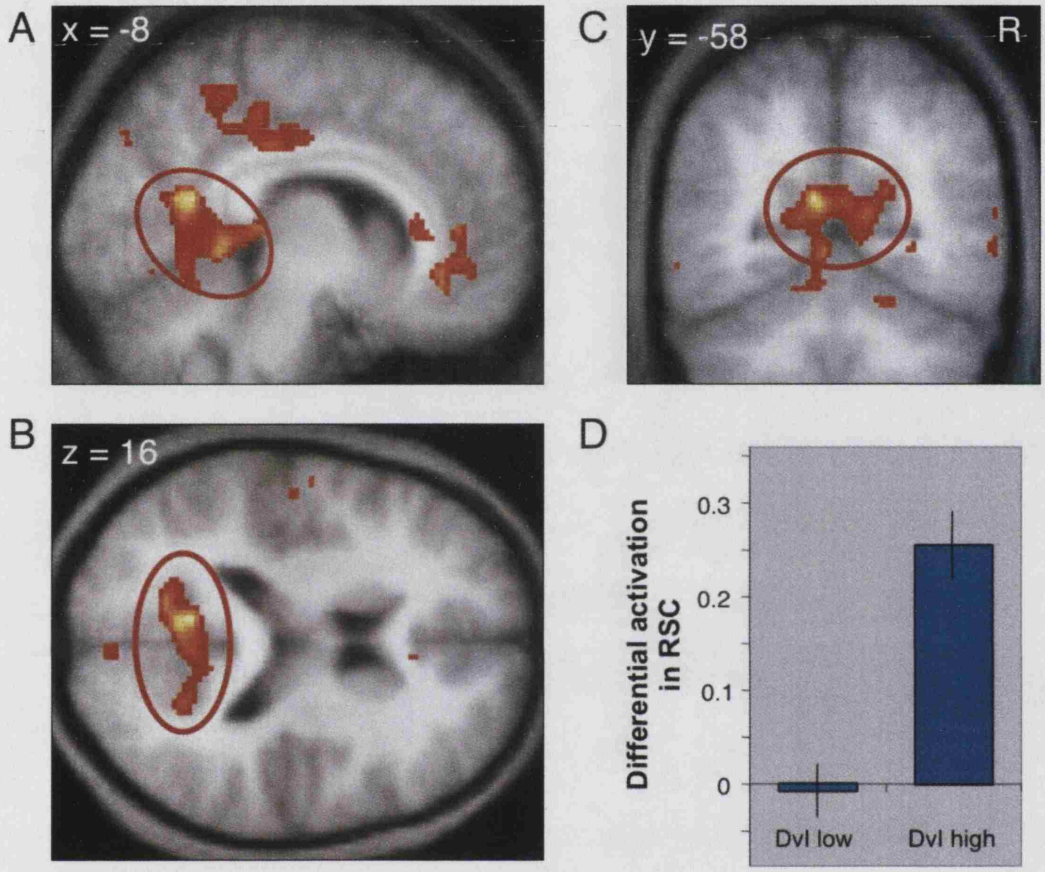

Figure 4.7: *Retrosplenial cortex shows interaction between task and intensity*

a-c. Sagittal, horizontal and coronal sections showing significant response during direct task compared to incidental task in retrosplenial / posterior cingulate cortex (RSC). However, this region was also highlighted as showing a significant interaction between task and intensity. Display as in Figure 4.3a.

d. Demonstration that main effect of task is a result of greater responses in RSC to high intensity expression faces during the direct task than the incidental task. y-axis represents size of differential response in peak voxel (% signal change). *"DvI low"* = direct versus incidental task for low intensity expressions only. *"DvI high"* = direct versus incidental task for high intensity expressions only.

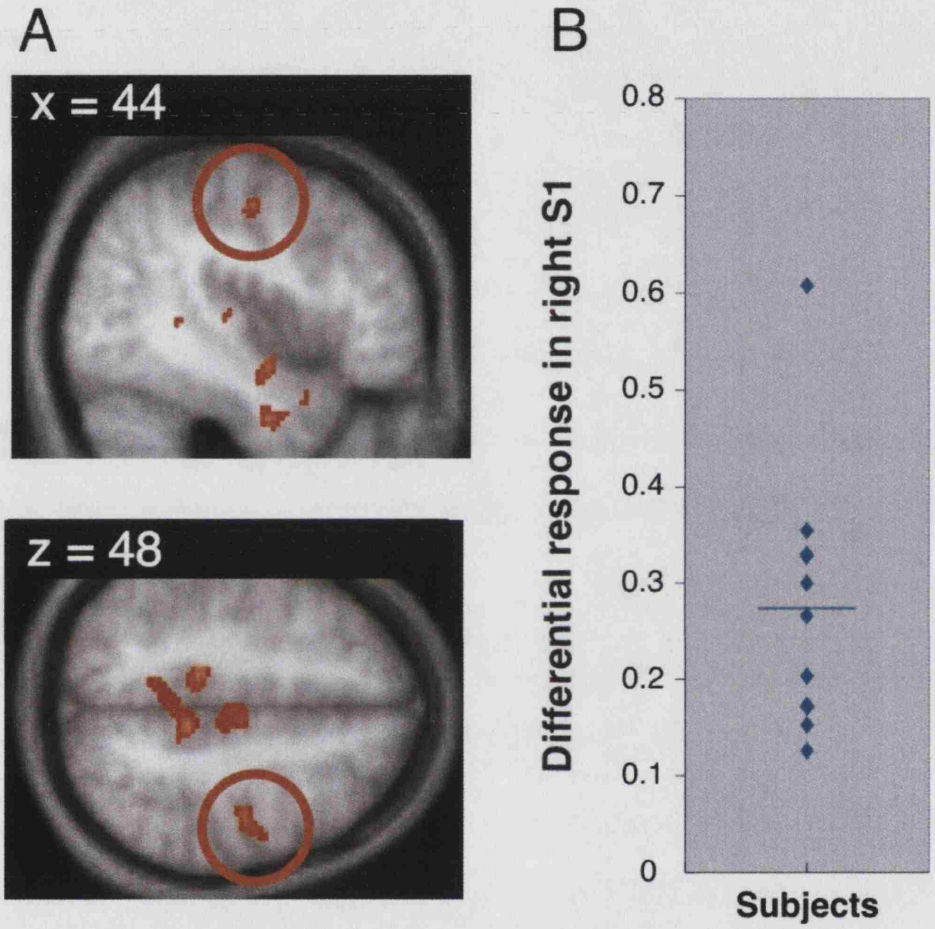

Figure 4.8: *Sensorimotor cortex responses in direct task*

a. Sagittal and horizontal sections showing significant response during direct task compared to incidental task in right sensorimotor cortex, including S1. Display as in Figure 4.3a.

b. Consistency of response in peak voxel in S1. Diamonds represent individual subjects, horizontal line, the mean response. y-axis represents size of differential response in S1 (% signal change).

Interactions between task and intensity

Only one region reached significance in the contrasts representing the interactions between task and intensity of emotion – posterior rostral motor cingulate (x,y,z=10,4,48, Z=5.18; $p<0.05$). A post-hoc test showed that this effect was driven by a greater response to low intensity faces in the direct condition. Although no other regions reached corrected significance, at an uncorrected threshold of $p<0.001$ voxels were detected in posterior cingulate close to the peaks of activation in the contrast of direct minus incidental task. This suggests that the main effects in this region were partly driven by interactions between intensity of stimulus and task. Post-hoc tests revealed that the main effect of task was predominantly a result of a greater response to high intensity emotional faces during the direct task (Figure 4.7d).

Differences between emotions

To examine for brain regions showing differential responses by emotion type random effects ANOVA models were used. Distinct contrasts were performed testing for the main effect of emotion, independent of task and intensity, for the interaction between emotion and intensity, for the interaction between emotion and task, and for the three-way interaction. Without specific hypotheses about interactions between emotion and task and the three-way interaction, these latter contrasts were only examined to ensure that in regions reported as significant for the main effects of task and interaction between task and intensity there was no interaction with emotion type.

No activations survived correction for multiple comparisons in the above contrasts for the differences between emotions. For completeness, I report descriptively activations in predicted regions at $p<0.001$ uncorrected. For the main effect of emotion, activation was seen in posterior right insula (x,y,z=42,-12,6, Z=3.73) where post hoc tests revealed that this effect reflected greater response to disgust faces. The interaction between emotion type and intensity revealed a greater response to high disgust faces in anterior cingulate (x,y,z=6,26,24; Z=4.53) and a decreased response to high intensity sad faces in anterior insula (x,y,z=40,18,-14; Z=3.83). Even at uncorrected thresholds of $p<0.05$ no evidence was found to support the hypothesis of a specific anterior insula response to disgust faces. No regions highlighted in the main effect of task or interaction between task and intensity were shown in the interactions between emotion and task and between emotion, intensity and task, indicating the independence of the main effects reported in these contrasts from interactions between the different emotions.

Despite the results of the conjunction analysis above (see "*Tests for similarities between emotions – conjunction analysis*" above) highlighting a common response within amygdala to high intensity expressions of the four emotions, there remains the possibility of an interaction between emotion type and intensity. For example, a high intensity expression of one emotion type might activate the amygdala to a greater degree than high intensity exemplars of other emotions, even though both activate amygdala relative to their low intensity counterparts. As the existence of such interactions might critically influence the interpretation of these results, such interactions were investigated at lowered statistical thresholds within the amygdala. Even at thresholds of $p<0.1$ uncorrected, there was no evidence for fear-selectivity in voxels within amygdala. Indeed, at such low thresholds, where false positives are

poorly protected against, the voxels which showed interactions by emotion type appeared more responsive to disgust or combinations of disgust and sadness than to fear.

Discussion

The goal of this study was to address the question of whether task engenders differential neural responses to the perception of facial expressions of emotion, and to what degree these neural responses are common to distinct categories of emotions. By using a multifactorial event-related fMRI design which required either direct or incidental processing of four different facial emotions the task-dependence of neural responses to these emotions was indexed as well as their specificity to one or more emotion type. A critical finding is the demonstration of task-independent amygdala responses in a conjunction of high intensity minus low intensity expressions across four basic emotions. Thus, the null hypothesis that at least one of these emotions failed to activate the amygdala can be rejected. The findings that the conjunction of high intensity minus low intensity across task and emotion yields amygdala activation, and an absence of interaction between intensity and task in the amygdala, provide evidence that this region responds to emotionally salient stimuli independent of task. This finding is consistent with previous indications that the amygdala shows a high degree of automaticity in responding to emotive stimuli, independent of task, attention or awareness (Morris et al., 1998b; Whalen et al., 1998; Vuilleumier et al., 2001; Dolan, 2002; Ohman, 2002; see also Chapters 6 and 7).

Although the neuroimaging and neuropsychological literature on the topic of perception of facial expressions of emotion has tended to emphasise roles for the amygdala in perception of negative facial emotions (Morris et al., 1996; Adolphs et al., 1999; Adolphs, 2002a), a compelling animal and human literature implicates the amygdala in positive affective learning and perception (Rolls, 1999; Johnsrude et al., 2000; Ono and Nishijo, 2000; Garavan et al., 2001; O'Doherty et al., 2001b; Parkinson et al., 2001; Baxter and Murray, 2002; Hamann et al., 2002). A more considered view, therefore, might be to suggest that the amygdala responds to stimuli of motivational significance, independent of their emotional valence. This will be discussed at greater length in Chapter 7 and the general discussion.

It is important to consider why previous imaging studies concerning perception of facial expressions have reported different results, particularly with regard to the role of the amygdala. The majority of previous studies have adopted block designs, which have two major disadvantages. The first is the inability of blocked designs to separate effects due to anticipation versus processing of a given stimulus. This is particularly important in the context of emotive stimuli where it is known that there exist specific substrates for prediction and that these are partially dissociable from those invoked by receipt (e.g. O'Doherty et al., 2002). The second is the possibility of differential habituation in emotion processing regions to repeated presentations of different emotions, an effect previously reported (e.g. Breiter et al., 1996; Phillips et al., 2001; Wright et al., 2001). Randomised event-related designs, as adopted here, are less susceptible to habituation effects. Thus, in previous block designs, amygdala responses may have differed between different emotions due to a role for this region in anticipation of a given emotional stimulus, or more rapid adaptation to one stimulus type than another. Either

of these effects might account for apparent discrepancy between the results of this study and those in the literature (e.g. Morris et al., 1996; Phillips et al., 1997, 1998) particularly with respect to the amygdala. Indeed, it should be noted that although the positive finding from a conjunction analysis allows the rejection of the null hypothesis that at least one of the high intensity expressions failed to activate amygdala more than its low intensity counterpart, it cannot be concluded that no differences exist in the amygdala between the processing of the distinct emotions. To draw such an inference would be to accept the null hypothesis tested when looking for differences between the emotions. It is noteworthy, however, that regions other than the amygdala (e.g. anterior cingulate and posterior insula) did show selective responses, which indicates that the experimental design had sufficient power to detect differential responses. It is possible that more events per condition or a higher field strength scanner might prove more sensitive to differences between conditions in this region. Although the peak voxel in right amygdala shows a trend towards greater responses to fearful faces (Figure 4.3), this trend does not even approach significance ($p>0.8$). Other voxels within the amygdala region highlighted by the conjunction do not show such a trend. As noted in the introduction to this chapter, the findings from human patients with lesions in the region of the amygdala are less consistent than assumed. A recent meta-analysis (Fine and Blair, 2000) highlights the fact that only two patients with fear-selective deficits (out of a total of thirteen) have been reported, and one other patient (EP) showed an anger-selective impairment. Although it is apparent that fear recognition is the most common impairment in this patient group, it is argued that fear is the hardest emotion for healthy subjects to recognise, and the association of fear recognition deficits and amygdala damage may partly reflect task difficulty effects (Rapcsak et al., 2000). Indeed, when the performance of neurologically intact subjects on the specific emotions

is factored into an analysis of amygdala-damaged patients, no fear-specific deficit is found (Rapcsak et al., 2000). It is interesting to note that single neuron recordings in human amygdala have failed to show a general selectivity for fear, instead finding single neurons with responses to single emotions, but no prevalence of one emotion type (Fried et al., 1997).

A contrast between the direct and incidental task revealed significant differential activity in rostral anterior cingulate and ventromedial prefrontal cortex, indicating these regions mediate explicit representation of emotions. Notably, ventral frontal cortex has previously been implicated in emotion recognition in faces and other modalities in neuropsychological studies (Hornak et al., 1996). This activation accords with a proposed role for ventromedial prefrontal cortex in intentional expression of emotional and social behaviour (Eslinger and Damasio, 1985; Grattan and Eslinger, 1992; Eslinger, 1998; Dimitrov et al., 1999; Anderson et al., 2003a). A mechanism for recognition of a conspecific's emotional states has been proposed to involve a form of simulation within one's own cortical circuitry (Adolphs et al., 2000; Adolphs, 2002b). In this regard it is noteworthy that anterior cingulate/ventromedial prefrontal cortex is involved in generation and reafferent mapping of bodily states (Critchley et al., 2000c) that constitute an integral part of emotional experience (James, 1884; Damasio, 1994). There is also evidence for enhanced activation in this region of ventromedial prefrontal cortex when subjects attend to their subjective emotional response to stimuli (Lane et al., 1997a).

A striking finding was enhanced activation in somatosensory cortex and insula during the direct task, though the latter failed to correct for multiple comparisons across

bilateral insula volume. Activation in these regions accords with the proposal that recognition of emotion involves simulation of the emotional display of a conspecific (Adolphs et al., 2000). However, it is interesting that activation in these regions is revealed in a main effect of task, and not in a main effect of intensity. The latter would be more compatible with a model in which simple perception of emotional faces entailed activation in these regions. The fact that the former is the case suggests that patients' deficits in emotion recognition tasks following lesions in ventral prefrontal and right somatosensory regions may reflect impairment in deliberative categorising of facial expressions despite intact basic perception. Activation in motor cingulate in the interaction between task and intensity of emotion, driven by an increased response to low intensity expressions during the direct task, might reflect greater effort for this condition. One possibility is that low intensity expressions required increased processing in the service of successful recognition.

The possible role of retrosplenial cortex in human emotional function has been highlighted by the suggestion that this region is commonly activated in neuroimaging studies of emotion (Maddock, 1999). This region is also strongly implicated in episodic memory retrieval processes (Fink et al., 1996; Maguire, 2001a, b). Even within the context of a role for this region in mediating familiarity, one recent study observed a trend suggestive of a specialisation for face stimuli over voice stimuli (Shah et al., 2001). In the current study, this region was activated by the main effect of task (direct versus incidental) and in the interaction between task and emotion intensity resulting from enhanced activity during presentation of high intensity faces specific to the direct task (Figure 4.7). Although a prediction from the literature on this region's responses to emotional stimuli would be a main effect of intensity, the data from the current

experiment suggest that it responds to emotionally-salient stimuli under explicit conditions. Indeed, it is noteworthy that five of the six "well-controlled studies" highlighted by a meta-analysis of functional imaging studies of emotion as showing activation in this region (Maddock, 1999) utilised tasks requiring explicit evaluation or recall of stimuli or situations. Thus, it seems that retrosplenial cortex is involved in direct evaluation of highly emotive stimuli which may reflect a necessity to recall stored representations of emotional information.

Both fusiform cortex and posterior superior temporal sulcus (STS) exhibit face-responsiveness as measured by fMRI (Kanwisher et al., 1997; Halgren et al., 1999; Hoffman and Haxby, 2000; Kesler-West et al., 2001). As discussed in the introduction to this thesis and in the previous chapter, a recently proposed neuroanatomical model of face perception highlighted dissociable roles of these two regions (Haxby et al., 2000). The model proposes that fusiform cortex is part of a core system for visual analysis of faces that responds to invariant aspects of faces such as identity, whereas STS codes changeable aspects of faces, including eye gaze and facial expression. The data presented in this chapter broadly conform to this model, in that STS responded more to high than low intensity expressions and to the emotional judgement relative to gender judgement. However, increased fusiform responses to emotional faces were also demonstrated, which might be attributed to greater processing resources being deployed to emotionally-salient stimuli (Dolan, 2002); effects that might relate to feedback from amygdala to visual areas (Amaral et al., 1992; Morris et al., 1998a). In STS, although the same general pattern of enhanced BOLD responses to emotional faces is seen, it seems probable on the basis of single unit studies in non-human primates (Hasselmo et al., 1989) and the fMRI data presented in Chapter 3 that this reflects specific coding for

facial expressions within this region, rather than a more general effect of enhanced attentional processing. These interpretations will be discussed in greater depth in the general discussion (Chapter 8). An additional dissociation between these two regions is the task-dependent nature of the response in STS that took the form of an additive effect of high intensity expressions and direct task, in the absence of an interaction. Attention to facial emotion (Streit et al., 1999; Narumoto et al., 2001), or even to emotive facial characteristics such as trustworthiness (see Chapter 6), activates STS. One possibility is that the direct task engenders the sort of "intention detection" that has been hypothesised to invoke this region (Frith and Frith, 1999; Jellema et al., 2000).

A provocative portion of the literature concentrates on highlighting dissociations between neural substrates for perception of different emotions (Phillips et al., 1998; Calder et al., 2001). In this context the paucity of regions displaying such dissociations is of interest. Although some differences were apparent at uncorrected thresholds in "limbic" areas (insula and anterior cingulate) these differences did not accord with predictions from the literature. For example, increased activation to disgust would be predicted in anterior insula (Phillips et al., 1997, 1998; Calder et al., 2001), and although some differences were evident in this region, they were not a result of selective responses to disgust faces. The lack of evidence for differences between emotions in this study might be construed as supportive of dimensional accounts of emotion, but these do not fit the data in a simple manner. For example, a dimensional account of emotion would not predict that a single region responds to high intensity exemplars of emotions that represent the opposite ends of the valence (e.g. happy and disgusted) and intensity (e.g. fearful and sad) spectra. Rather, the behavioural evidence for categorical emotion perception is persuasive (e.g. Etcoff and Magee, 1992; Young et

al., 1997) but the coding for different emotions may exist on a spatial scale that borders the limits of the current technique of subtractive fMRI. This could conceivably take the form of single neurons coding for distinct emotions within the same brain region (e.g. STS). See the experiment in Chapter 3 for an attempt to use a different neuroimaging approach to find brain regions coding for multiple emotions.

Summary

The data presented in this chapter indicate a common neural substrate for perception of different facial emotions involving the amygdala, extrastriate and fusiform cortex and STS. Taken in conjunction with the failure to demonstrate regions exhibiting convincing differences between the four different emotions studied here, these data challenge the concept of functional segregation for perception of individual emotions at a gross anatomical level. Increased responses to highly emotional faces in amygdala and fusiform were task-independent, concomitant with a hypothesis of obligatory processing of emotion in these regions. Task-dependent increases in neural responses in ventromedial prefrontal cortex and somatosensory regions suggest that the impairment in emotion recognition of patients with lesions in these regions is a function of lack of integration between intact perceptual systems and systems that provide "somatic" cues that can bias overt emotional judgements.

Chapter 5: Spatial frequencies and fearful face processing

Introduction

The two previous chapters described experiments designed to identify brain regions responsive to emotional expressions in faces. In this chapter, I describe an experiment designed to explore the fundamental visual components of fearful faces that engage emotional systems. Specifically, I report a study that addresses the roles of different spatial frequency bands in perception of, and the brain's response to, fearful faces.

Schyns and colleagues have described the influence of different spatial frequency components in face perception using hybrid stimuli composed of two overlapped faces, one high-pass and a second low-pass filtered (Schyns and Oliva, 1999; Morrison and Schyns, 2001). When rapidly presented, subjects report seeing only a single face but can be directed to report one or the other component of these hybrids depending on task instructions. For example, gender judgement results in an approximately equal use of each spatial scale (Schyns and Oliva, 1999). By contrast, classifying faces as emotional or non-emotional biases subjects to use high spatial frequencies (HSF), whereas classifying the distinct type of emotion biases subjects to low spatial frequencies (LSF).

Functional neuroimaging experiments, building on previous behavioural and psychophysiological work (Esteves and Ohman, 1993; Esteves et al., 1994), have demonstrated that faces with emotional significance may evoke selective neural responses even when conscious awareness is prevented by backward masking

techniques (Morris et al., 1998b; Whalen et al., 1998). Analysis of effective connectivity suggested an enhanced interaction between amygdala and a colliculo-pulvinar (extra-geniculostriate) pathway during processing of masked fear-conditioned stimuli (Morris et al., 1999). This is supported by studies of a blindsight patient (GY), who showed both amygdala and superior colliculus activation to fearful or fear-conditioned faces presented within his scotoma (Morris et al., 2001b). Amygdala and orbitofrontal cortex and fusiform regions were also activated by emotional stimuli extinguished from awareness in a neglect patient (Vuilleumier et al., 2002).

It has recently been demonstrated that an amygdala response to fearful expressions in faces is preferentially mediated by visual information carried in the LSF domain (Vuilleumier et al., 2003a). Such LSF information might be conveyed through subcortical pathways in superior colliculus and pulvinar under conditions where conscious vision is suppressed (Rafal et al., 1990; Dodds et al., 2002; Sahraie et al., 2002). In the study described in this chapter, the differential roles of high and low frequency information processing channels on neuronal and behavioural responses to faces are further addressed. In addition, the degree to which responses to these channels depend upon explicit perception is investigated. Using hybrid stimuli, I addressed the question of whether emotionality in an LSF channel can influence processing in an HSF channel. This manipulation also allowed me to explore the degree to which such influences are independent of the visual channel currently perceived. In an event-related fMRI design, participants made gender judgements on hybrid faces. Each stimulus contained a male and a female face, one in high and one in low spatial frequencies, while the expression of each face also varied independently. Subjects used the HSF and LSF channels flexibly to make gender judgements, so that effects of both

low and high spatial frequency components could be examined independently, either when perceived or when not reported. As faces in either frequency band could be fearful or neutral, this resulted in a 2 (reported spatial frequency) X 2 (emotion at LSF) X 2 (emotion at HSF) design (Figure 5.1b).

Methods

Stimuli

The stimuli were derived from 12 individuals (six males) from the Karolinska Directed Emotional Faces (KDEF) set (Lundqvist and Litton, 1998), chosen on the basis of having similar face shape and size. Neutral and fearful expressions were selected from all identities. Hybrid stimuli were generated according to a method outlined by (Schyns and Oliva, 1994). Briefly, a face of one gender was low-pass filtered and combined with a high-pass filtered image of a face of the opposite gender. The low pass filter was set at 6 cycles per face and the high-pass filter at 24 cycles per face. These values are similar to those used in a study using non-hybridised filtered stimuli (Vuilleumier et al., 2003a) and within the range expected to produce flexible gender categorisation (Schyns and Oliva, 1999). When presented in the scanner, these values corresponded to approximate cut-offs of 1.5 and 6 cycles per degree of visual angle, again similar to previous studies. Faces at each spatial frequency scale could be either fearful or neutral, but were always of opposite gender from one another. Thus, a total of 288 stimuli were generated (12 identities, 6 male/6 female; each crossed with the six of the opposite gender, with four possible pairings of emotional expressions: neutral/neutral,

fearful/neutral, neutral/fearful, fearful/fearful). See Figure 5.1 for example stimuli and experimental design.

Subjects

Fourteen subjects (eight females) agreed to take part in the study. Data from one female subject were excluded due to excessive head movement (greater than one voxel diameter). No subject reported perceiving hybrid stimuli in debriefing. The age range of included subjects was 22-44 (mean 30). All were right-handed, free of significant neurological and psychiatric medical history, and had normal or corrected-to-normal vision.

Experimental task

Subjects made gender discriminations on hybrid faces presented for 90ms and minimum stimulus onset asynchrony was 3s. 90ms was chosen after pilot experiments with the stimuli and visual presentation apparatus suggested that such a presentation time ensured an equal likelihood of reporting high vs. low spatial frequency stimuli within cross-gender face hybrids, as reported by Schyns and Oliva (1999). In addition to the 288 stimuli, 62 null events were included, with a total session length of approximately eighteen minutes.

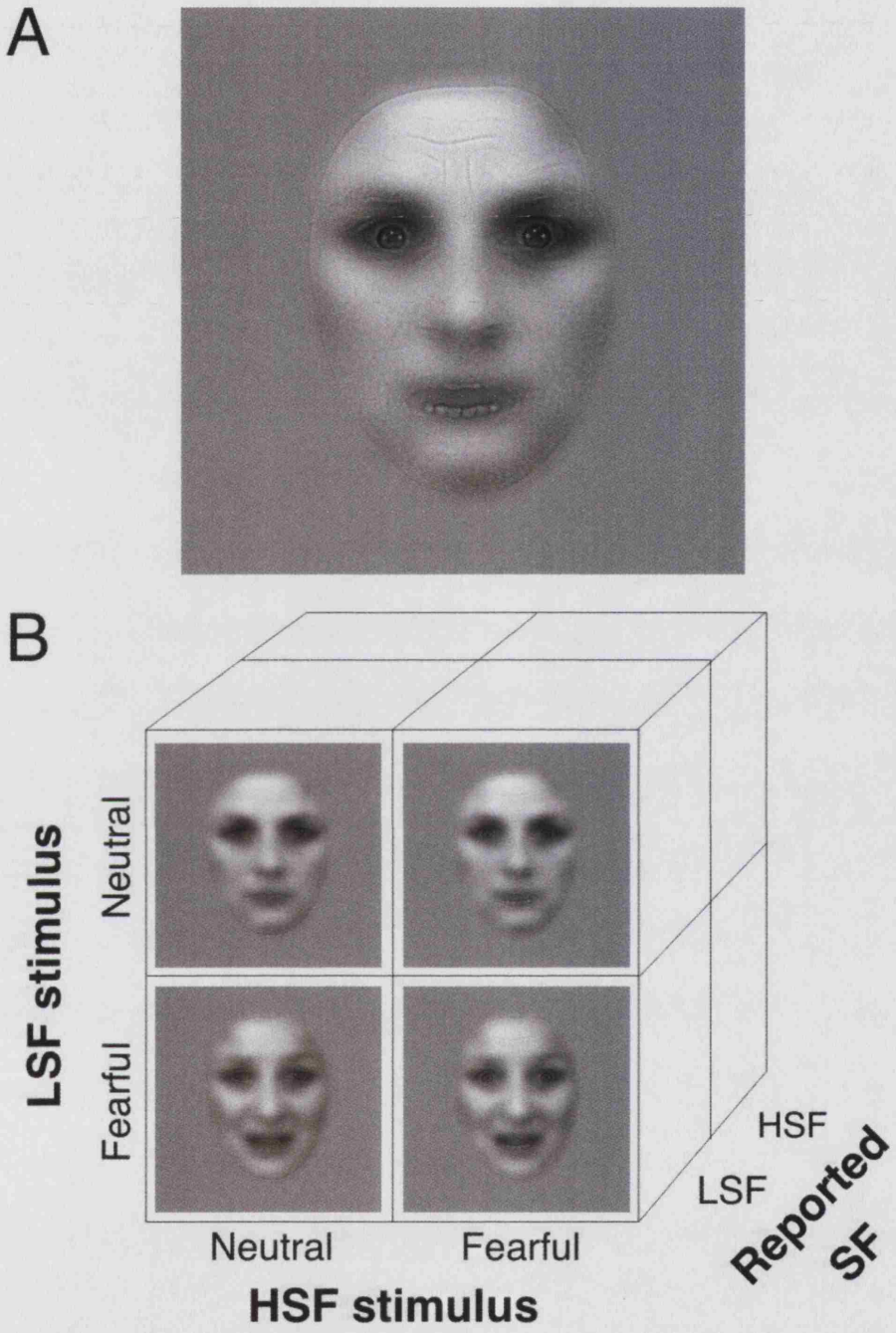

Figure 5.1: *Example stimuli and experimental design*

a. Example stimulus with neutral female at low spatial frequencies and fearful male at high spatial frequencies. Perception is typically dominated by HSF cues when looking at such stimuli non-tachistoscopically. To see the LSF, squint or increase the viewing distance.

b. Experimental design: subjects could report either the low or high SF. At either SF the face could be fearful or neutral, resulting in a 2 X 2 X 2 factorial design.

Subsequent to the main experimental session, whilst still in the scanner, subjects made gender discriminations on individual faces presented in either high or low frequencies, filtered with the same parameters as the component frequencies as the hybrid faces. Gender judgements from this session served to interpret subjects' classification responses from the hybrid sessions.

fMRI scanning

A Siemens VISION system at 2 Tesla was used to acquire gradient-echo, echoplanar T2*-weighted images with BOLD contrast. Volumes comprised 33 X 3.3mm axial scans with 3mm in-plane resolution and were continuously acquired every 2.5s. In total, 435 volumes were acquired for each subject and the first five volumes of each run discarded. Additionally a T1-weighted structural image was acquired in each subject. These structural images were normalised into a standard space using the normalisation parameters applied to the EPIs and subsequently averaged for overlay of statistical parametric maps.

Data analysis

Initial pre-processing, including realignment and slice-time correction, of imaging data was undertaken using SPM99. Subsequent pre-processing, including normalisation and smoothing, and data analysis was undertaken using SPM2b.

For event-related analysis, 12 event types were modelled for each subject. These corresponded to the four possible combinations of emotional expression in the high and

low spatial frequency channels (HSF neutral/LSF neutral, HSF neutral/LSF fear, HSF fear/LSF neutral, HSF fear/LSF fear) crossed with three levels of report by the subject (HSF reported, LSF reported, or ambiguous). Ambiguous reports resulted from absent button presses and trials in which both identities comprising the hybrid were labelled as the same gender during the debriefing session in the scanner (range 9-55%; mean=43%). Regressors were constructed for each event type using a synthetic HRF and its temporal derivative, as well as modelling interactions between event and experimental time. The generalised linear model was used to estimate effect sizes for the different events types, as described in the general methods chapter. Here I report results only from contrasts using time-independent effects characterised by the HRF alone.

Given the strong *a priori* predictions from Vuilleumier et al. (2003a) a statistical threshold of $p<0.001$ uncorrected for multiple comparisons with an extent threshold of $100mm^3$ (> 4 contiguous voxels) was used to define activations. In regions predicted *a priori* (fusiform cortex), results that survive a correction for multiple comparisons across a small volume of interest (Worsley et al., 1996) are reported. A map from a random effects analysis of faces versus scrambled faces (thresholded at $p<0.001$ uncorrected) from an independent subject group was used to define face-responsive fusiform cortex for small volume correction. In addition, peak voxels from previous results with filtered faces (Vuilleumier et al., 2003a) were inspected at reduced thresholds to ascertain whether activation was seen in critical voxels identified in this previous study at $p<0.05$ uncorrected (there being no multiple comparisons problem with a single predicted voxel).

Results

Behavioural

All subjects' reports from the spatial scales tended to be biased towards the LSF scale, as opposed to the unbiased reporting of gender demonstrated by (Schyns and Oliva, 1999). This discrepancy might be explained by the different stimuli used here (fearful [rather than happy and angry] faces from a different database) or some specific feature of the protocol (e.g. slightly longer presentation time, exact visual angle subtended by the stimuli). In any case, the bias to LSF was small, with proportions ranging from 50% to 70% reports from LSF, and mean 60.5% bias. Reaction times (RTs) were analysed by repeated measures ANOVA with SPSS (SPSS inc, Chicago). There was a main effect of subjects' report on reaction time, with reporting HSF associated with slower RTs (LSF reported: 839±48ms, HSF reported: 931±54ms (means±SEM); $F_{(1,12)}=12.1$; $p<0.01$). The interaction between reported SF and LSF fear was further explored, testing for a reaction time change in reporting high SF faces in the context of fear in LSF (Figure 5.3a). Only a trend to slower responses was evident ($F_{(1,12)}=2.1$; $p=0.18$). No other main effect or interaction approached significance (all $p>0.25$).

Neuroimaging

Main effect of report

Whereas no differential fMRI responses were seen in regions of interest (e.g. fusiform, pulvinar or superior colliculus) as a function of the spatial frequency range reported during gender judgements, greater responses when reporting faces from LSF

components were seen in posterior cingulate and marginal segment of cingulate sulcus (x,y,z=-9,-21,-33, Z=4.12; x,y,z=-12,-45,63, Z=3.93), lateral orbitofrontal cortex (x,y,z=-39,33,-24; Z=3.93), superior frontal gyrus (x,y,z=-30,24,36; Z=3.73) and entorhinal cortex (x,y,z=-18,-21,-27; Z=3.73). No cortical regions showed an opposite effect.

Main effect of LSF fear

A main effect of fear in the low spatial frequency components of faces (compared to LSF neutral) was found in right fusiform cortex (x,y,z=39,-48,-27; Z=4.33; $p<0.05$ SVC; Figure 5.2). This effect was independent of the reported spatial frequency channel, as confirmed by a conjunction analysis across the simple effects of LSF fear (relative to neutral) when the LSF channel was reported and when the HSF channel was reported. This indicates that fear in the LSF channel caused greater activation in right fusiform cortex regardless of which SF channel the subject reported.

While behaviourally the overall reaction time interaction between LSF emotion and reported SF was non-significant (see the section above describing behavioural results), a significant correlation was evident between the RT measure of this interaction for each subject and the magnitude of activation to LSF fear in the identified peak in right fusiform ($p<0.05$; Figure 3b). In other words, the degree to which a subject showed enhanced fusiform activity to LSF fear relative to LSF neutral (as a main effect) predicted an RT cost during explicit perceptual report of the HSF faces.

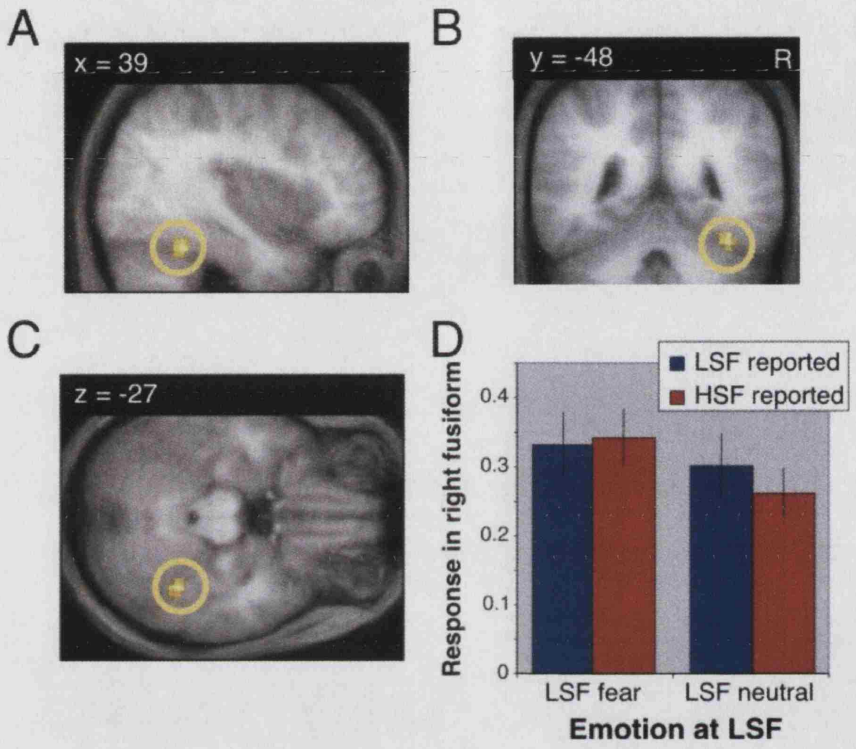

Figure 5.2: *Fusiform response to fear in low spatial frequencies*

a-c. Sagittal, coronal and horizontal sections of SPMs showing response in right fusiform cortex in the contrast of main effect of LSF fear (LSF fear – LSF neutral). Peak at x,y,z=39,-48,-27; Z=4.33; $p<0.05$ SVC. Activation displayed at $p<0.001$ uncorrected on mean structural image from the group of subjects.

d. Response estimates from peak fusiform voxel showing response to LSF fear is greater than to LSF neutral independent of subjects' report (i.e. whether they report perceiving LSF or HSF stimulus). Height represents % signal change, error bars are standard error.

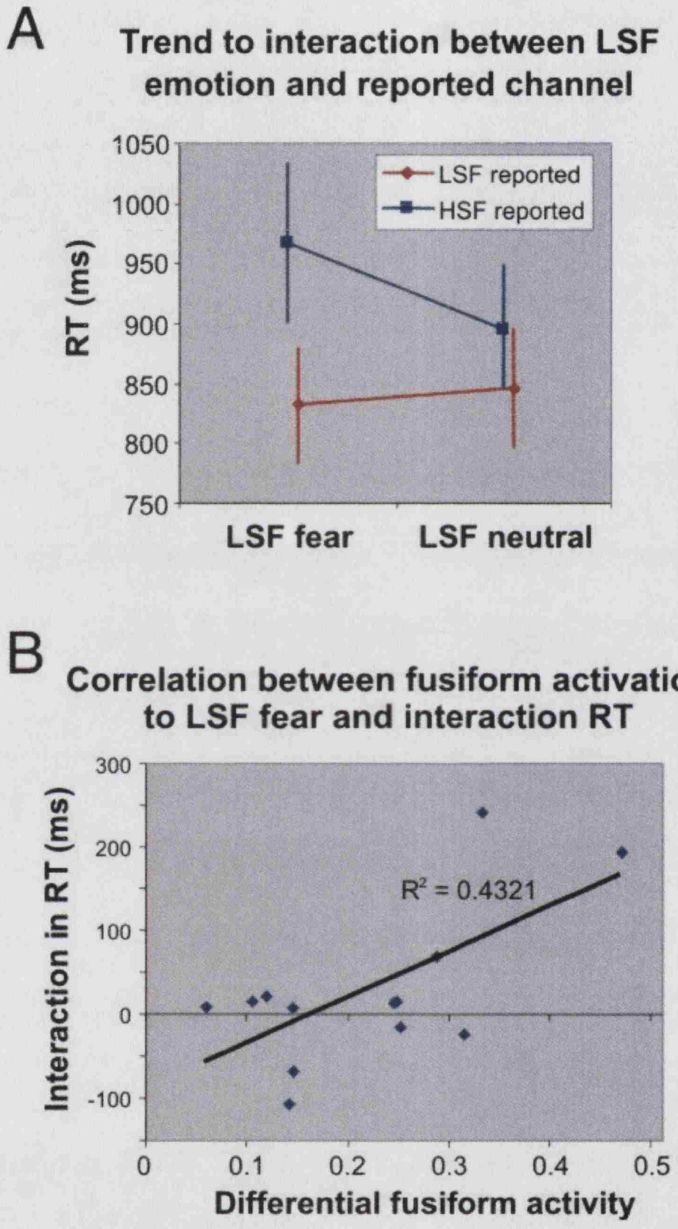

Figure 5.3: *Behavioural effect of fear in LSF*

a. Trend level (p=0.18) interaction between fear in LSF and subjects' report, suggesting slower RTs when fear is present in LSF and subject reports HSF stimulus

b. Significant correlation between fusiform activity to main effect of LSF fear and behavioural interaction between report and LSF fear in reaction times ($p<0.05$). This indicates that greater fusiform activity to LSF fear stimuli relates to a behavioural slowing when responding to the HSF components of the stimulus.

Responses in specific regions at reduced thresholds were additionally examined for, based upon *a priori* hypotheses. In bilateral amygdala/peri-amygdaloid cortex, weak activation was evident for the main effect of LSF fear (x,y,z=-18,-6,-33, Z=1.78; x,y,z=33,-6,-24, Z=1.81), though not at exactly the same peaks as previous work (for comparison: x,y,z=-20,–10,–28 and x,y,z=20,–10,-30, Vuilleumier et al., 2003a). In lateral postero-inferior thalamus, weak activation was also evident in a peak voxel found in previous work (x,y,z=9,-21,-9; Z=1.67; Vuilleumier et al., 2003a). As noted by Vuilleumier et al. (2003a), this activation extended broadly throughout posterior and lateral thalamus and into superior colliculus (x,y,z=0,-30,-3, Z=2.09).

Main effect of HSF fear

The peaks reported for LSF fear were additionally checked for a main effect of HSF fear. No evidence was found for effects of HSF fear in these voxels (fusiform peak, Z=1.00; postero-lateral thalamus, Z=-0.94; left amygdala, Z=-1.25; right amygdala, Z=0.09; all $p > 0.15$). This replicates a lack of significant responses to fearful expression conveyed by HSF cues in faces in the same regions as LSF fear (Vuilleumier et al., 2003a), though there were responses in posterior cingulate (x,y,z=15,-36,39; Z=4.02), motor cortex (x,y,z=42,0,36; Z=3.75), medial prefrontal cortex (x,y,z=9,63,30; Z=3.63) and lateral orbitofrontal cortex (x,y,z=45,35,-21; Z=3.53).

Interaction of LSF emotion and reported SF

Finally, the interaction of the factorial design was tested, i.e. differential effects of LSF fear when the LSF face was reported as compared to when the HSF face was reported. Such an interaction between LSF emotion and reported SF was observed in right orbitofrontal cortex (x,y,z=24,42,-15, Z=3.42; Figure 5.4), reflecting a greater response in this region to fearful expression in LSF only when subjects report the fearful face in this LSF channel (Figure 4b; simple effect of LSF fear when LSF reported, Z=3.43; simple effect of LSF fear when HSF fear reported, Z= -0.82). This effect was independent of the expression in the concomitant HSF face (i.e. whether it was fearful or neutral).

Discussion

In this experiment I tested the hypothesis of preferential neural responses to fearful faces in low spatial frequency components of an image, and a hypothesis of an emotion-dependent modulation of face-responsive region of fusiform cortex. Using rapidly-presented hybrid faces, in a task known to induce flexible use of spatial frequency components (Schyns and Oliva, 1999), it was possible to distinguish between brain regions responding to fearful faces in one spatial frequency band independent of whether or not gender information from this set of spatial frequencies was reported. The principal finding was an enhancement of fusiform responses to faces containing fear in LSF components, with this effect being independent of whether subjects actually report seeing the low or high spatial frequency component of a hybrid stimulus.

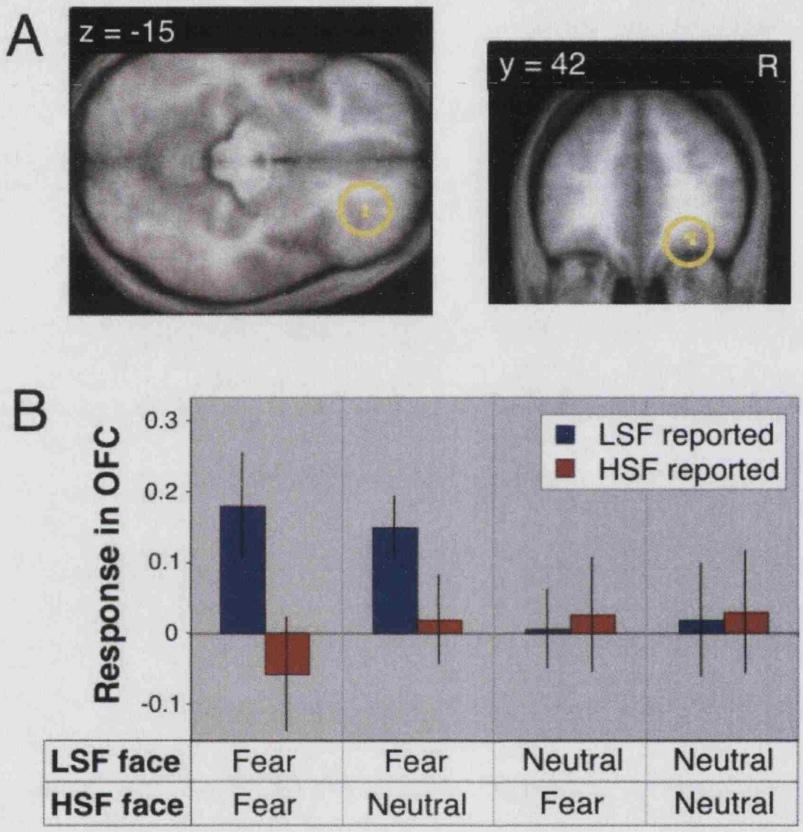

Figure 5.4: *Orbitofrontal cortex shows an interaction between report and LSF fear*

a. SPM showing response in right orbitofrontal cortex in the interaction between LSF fear and behavioural report. Peak at x,y,z=24,42,-15; Z=3.41; $p<0.001$ uncorrected. Display as in Figure 5.2a.

b. Parameter estimates (units: % signal change; error bars are standard error) from peak voxel in orbitofrontal cortex showing response to LSF fear is greater than to LSF neutral only when subjects report LSF components of the stimulus.

This result extends previous findings concerning the responsivity of fusiform cortex to emotionality in a face. It has repeatedly been demonstrated that fusiform cortex shows greater activation to emotional compared to neutral faces (Breiter et al., 1996; Dolan et al., 1996; Morris et al., 1998a; Vuilleumier et al., 2001; Pessoa et al., 2002; Surguladze et al., 2003; see also Chapter 4). A recent report suggested that despite greater responses in fusiform cortex to HSF components of faces (regardless of expression), fearful LSF faces activate this region more than neutral LSF faces. Conversely, fearful HSF faces do not activate fusiform more than neutral HSF faces (Vuilleumier et al., 2003a). The current data indicate that even in the context of hybrid stimuli, subjects reporting HSF components show modulation of fusiform activity by unreported fearful LSF components.

It has been suggested that this fusiform modulation by emotionality is subserved by feedback from amygdala (Morris et al., 1998a; Dolan, 2002). Thus, an obligatory fear-detecting process engaging the amygdala may modulate, through re-entrant feedback, fusiform responses to a current stimulus. It has also been suggested that both thalamic (pulvinar) and amygdala responses are selectively tuned to extract LSF cues in faces (Vuilleumier et al., 2003a). This suggests that perceptual processing within an HSF channel, thought to preferentially drive fusiform activity (Vuilleumier et al., 2003a), might nonetheless be affected by emotionality in an LSF face. In the current study evidence was found for such a behavioural effect in individual reaction time measures, indicating that the interaction between emotion in the LSF channel and reported SF was significantly correlated with strength of differential activation (LSF fear > LSF neutral) in right fusiform. This indicates a strong coupling between the modulation of fusiform by emotion and the efficiency of perceptual processing of a HSF stimulus.

Vuilleumier et al. (2003a) have demonstrated enhanced activation to fear within low-pass filtered faces in superior colliculus and pulvinar, in addition to the amygdala. Conversely, there was no evidence for enhanced activation to fear expression in high-pass filtered faces in these regions or fusiform cortex. Consistent with this, in the current study greater colliculo-thalamo-amygdala activation was found to LSF representations of fearful faces. Although this activation was observed only at low threshold, it was found for voxels in specific predicted regions, reducing the multiple comparisons problem that usually complicates whole-brain statistical analysis. Importantly, these effects were also independent of the explicitly reported SF channel.

Although the previously demonstrated effect of LSF fear in activating the amygdala (Vuilleumier et al., 2003a) was replicated here, the strength of this effect was weaker. One possible reason is that the hybrid stimuli were all unusual relative to conventional faces and non-hybrid filtered faces, and thus the amygdala may have been more active in "neutral" conditions than in other studies, an explanation consistent with the concept that the amygdala is activated in directing attention to ambiguous stimuli (Whalen, 1998; Davis and Whalen, 2001). Interestingly, the peak voxel from the left peri-amygdaloid focus showed greater activity in all eight conditions than baseline. Additionally, the amygdala is known to habituate in response to emotional stimuli (Breiter et al., 1996; Buchel et al., 1999; Phillips et al., 2001; Wright et al., 2001), and the current experiment, at 18 minutes, is longer than other similar studies. Indeed, there was evidence for linear habituation effects in amygdala with time (x,y,z=-21,0,-33, Z=2.37; x,y,z=-27,-6,-24; Z=2.05).

In contrast to fusiform, which showed effects of LSF fear independent of subjects' report, right orbitofrontal cortex exhibited an interaction between LSF fear and reported SF. In this region an enhanced response to fear conveyed by the LSF component occurred exclusively when this LSF component was reported (Figure 4b). This finding accords with the view that sectors of ventral prefrontal cortex are specifically involved in conscious, deliberative emotional processing (Damasio, 1994). It is possible that processing of LSF fear stimuli will cause a downstream effect in OFC only if the LSF stimulus is perceived. This interpretation converges with recent fMRI findings from a patient with parietal neglect (Vuilleumier et al., 2002) where a very similar peak in right OFC (the adjacent voxel) showed greater responses to fearful than neutral faces only when the face was consciously perceived as opposed to neglected. It is interesting to also note that recent human electrophysiological work has identified an early (~120ms) frontal ERP component that distinguishes fearful faces from neutral (Eimer and Holmes, 2002) but is abolished when spatial attention is diverted from the fearful face (Holmes et al., 2003). This parallels the demonstration of orbitofrontal responses to LSF fear stimuli only when the LSF is the attended frequency band.

Summary

Using a novel experimental design I have demonstrated that emotional information in faces contained within low spatial frequencies influence fusiform responses independently of whether the gender corresponding to those spatial frequencies is reported or not. The magnitude of this fusiform modulation predicts the degree of a behavioural slowing on RTs when reporting the gender of a high spatial frequency face, as a function of the (independently-manipulated and task-irrelevant) emotion present in

low spatial frequencies. In contrast to fusiform, and concordant with neuroimaging and electrophysiological data of attention-dependent frontal responses to fearful faces (Vuilleumier et al., 2002; Holmes et al., 2003), orbitofrontal cortex responds to LSF fear cues in faces only when LSF is reported. These data confirm that emotional responses may occur partly independent of conscious visual processing of HSF within geniculostriate and ventral temporal regions that are critical in face perception, whereas frontal responses may specifically reflect the conscious emotional percept.

Chapter 6: Processing of trustworthiness from faces

Introduction

The previous experiments described in this thesis have characterised a set of neural systems for processing emotional faces, specifically demonstrating that some regions respond automatically to emotional expressions in contrast to others that engage in a task-dependent manner. These regions are to some degree dissociable from those engaged by facial identity processing. In this chapter, the brain's response to a facial characteristic more complex than emotional expression is explored – that of trustworthiness.

It has been conjectured that human survival has to a large extent depended upon accurate social judgements and that an evolutionary consequence of this are modular cognitive processes devoted to these functions (Humphrey, 1983). Neuropsychological studies and human functional imaging have provided partial support for this idea of a dedicated 'social intelligence', particularly studies that address perception of facial expression (e.g. Adolphs et al., 1994; Hornak et al., 1996; Morris et al., 1996; Phillips et al., 1997). However, facial emotional expression is only one aspect of social judgement made about others. In many situations, individuals must also decide whether another person is someone to approach or avoid, trust or distrust. Preliminary evidence regarding the neural underpinnings of this sort of evaluative judgement comes from studies of patients with brain lesions which have demonstrated that patients with bilateral amygdala lesions make abnormal social judgements about others based on facial appearance (Adolphs et al., 1998). These abnormalities are most pronounced in

relation to faces which received the most negative ratings by control subjects. Notably, such deficits are not apparent in subjects with unilateral amygdala lesions (Adolphs et al., 1998). It has also been reported that patients with damage to ventromedial prefrontal cortex have difficulties with trustworthiness decisions (Eslinger and Damasio, 1985; Damasio, 1994).

As described in the introduction to this thesis (Chapter 1) the most influential neurobiological model of social cognition is that of Brothers (1990). Drawing inferences largely from neurophysiological recordings in non-human primates, this author postulated that the superior temporal sulcus acts as association cortex for processing conspecifics' behaviour. Socially relevant information is subsequently labelled by emotional systems located in amygdala and orbitofrontal cortex. More recent models of human social cognition also include sensory regions such as the face-processing area in fusiform gyrus and somatosensory cortex (including insula, SI and SII, Adolphs, 1999, 2001, 2003b).

In the study described in this chapter I used event-related functional magnetic resonance imaging (fMRI) to ascertain the neural substrates mediating evaluative social judgement. There is evidence that processing of facial emotion can be implicit, occurring when subjects make judgements about facial attributes unrelated to emotion (e.g. Morris et al., 1996; Phillips et al., 1997; Blair et al., 1999, see also Chapters 3-5). To establish whether trustworthiness judgements might be similarly processed, a paradigm in which subjects viewed faces either while making explicit judgements whether or not an individual was trustworthy or when making an unrelated age assessment was used. To account for individual differences in trustworthiness

judgement, ratings of trustworthiness for each stimulus from each subject were acquired post-scanning and these ratings used as parametric covariates in the subsequent analysis. Based upon models of social cognition (Brothers, 1990; Adolphs, 1999, 2001, 2003b; Allison et al., 2000) along with the neuropsychological findings (Adolphs et al., 1998) amygdala, orbitofrontal cortex, fusiform gyrus and superior temporal sulcus formed regions of interest in the statistical analysis.

Methods

Stimuli

120 greyscale frontal images of Caucasian male faces were selected from a larger selection of images following a pilot study outside the scanner. The images were selected to cover a range of trustworthiness scores rated by the subjects in the pilot study (n=30; 13 females, 17 males; age range: 17-32; mean age 23.5), but score as low as possible on ratings of 'happiness' and 'anger'. Gaze and head direction of all stimuli was directly forwards. Stimuli were adjusted to be of approximately equal size and luminance and manipulated such that each face was centred on a grey background in a 400 X 400 pixel image. Of the 120 stimuli used in the imaging study, 60 were high school student photographs and 60 photographs of university students. There was no significant difference in average trustworthiness score between photos of schoolboys and photos of university students (Mann-Whitney U test, $p>0.90$).

Subjects

Informed consent to partake in the study was obtained from 16 right-handed Caucasian volunteers (8 male, 8 female; age range 18–30 years; mean age 23.3 years). Two subjects (both females) were excluded from the analysis; one revealed psychiatric history after scanning, another provided extreme trustworthiness ratings, appearing to invert the ratings scale (see 'Debriefing' below; Spearman's rho of correlation of ratings with mean of all other subjects=-0.445; for all remaining subjects Spearman's rho values were >0.3). All remaining subjects were free from psychiatric or neurological history.

Psychological task

The scanning session for each participant was divided into two parts. In one half of the session 60 faces were presented sequentially and participants made a judgement, indicating with a push-button response, whether the face was a high school or university student. In the other half of the session they judged whether the face was trustworthy or untrustworthy. The order of tasks was counterbalanced between participants. At the start of each task a word appeared on screen informing the subject of the task requirement ("School/Uni" or "Trustworthiness").

Stimuli were presented on a grey background once each in random order randomly interspersed with 60 null events. Each stimulus was presented for 1s with an inter-trial interval of 2s. Between faces a fixation cross was presented. Null events were of 3s

duration during which time a fixation cross remained on screen. Stimuli subtended visual angles of approximately 10° vertically and 5° horizontally.

Image acquisition

Subjects were scanned during task performance using a 2T Siemens VISION system to acquire BOLD-weighted EPIs. Each volume comprised 33 X 2.2mm axial scans with 3-mm in-plane resolution and volumes were continuously acquired every 2.5s. Each run began with 5 'dummy' volumes (subsequently discarded) to allow for T1 equilibration effects. Additionally a T1-weighted structural image was acquired in each subject (with one exception).

All functional volumes were realigned, slice timing corrected, normalised into a standard space and smoothed with an 8-mm FWHM Gaussian kernel using the methods described in the general methods chapter (Chapter 2).

Debriefing

After scanning, participants undertook a self-paced task in which they rated all the faces on a scale of trustworthiness from 1 (highly untrustworthy) to 7 (highly trustworthy). When all 120 faces had been rated, a second task was performed, in which participants named emotions that they perceived in the faces by means of a seven way forced choice procedure (neutral, happy, sad, angry, disgust, fear, surprise). In order to assist subjects with this task they were given a printed sheet with photographs of one face from the Ekman and Friesen (1975) series expressing each of these seven emotions.

Emotion ratings for stimuli

An additional set of 16 subjects (10 males, 6 females; age range 19-34 years; mean age 24 years) undertook a task in which they rated the degree of emotional expression within each face on each of four basic emotions (anger, fear, happiness and sadness) in turn. Ratings were from 1 (neutral for this particular emotion) to 7 (highest degree of this particular emotion). Participants rated each face on a given emotion and then each face on a different emotion until all faces had been rated for all emotions (with order of emotion judgements and stimulus order within block both randomised).

Data analysis

Imaging data were analysed with SPM99. The experimental design allowed a parametric factorial analysis whereby trustworthiness comprised a parametric regressor and task (age or trustworthiness judgement) the second factor.

The presentation of each face was modelled using the canonical HRF and its temporal derivative, with parametric modulators to model subject-specific trustworthiness judgements (see general methods). Random effects analysis across the 14 subjects was used for statistical inference. In regions of interest, a correction for multiple comparisons across a small volume of interest was applied to the ensuing p-values in this region (Worsley et al., 1996). Activations in predicted regions surviving this correction at $p<0.05$ are reported as small volume corrected (SVC). Volumes of interest for amygdala, orbitofrontal cortex and for STS were defined by drawing a mask around the regions bilaterally on a normalised T1 structural image with reference to an atlas of

human neuroanatomy (Duvernoy, 1999) using the software package MRIcro (http://www.mricro.com). Total volume of the amygdala mask was approximately $10cm^3$, volume of the orbitofrontal mask approximately $50cm^3$ and volume of the STS mask approximately $20cm^3$. In the case of the fusiform gyri, small volume correction was based upon a sphere of 10mm radius centred upon coordinates derived from a previous study (Vuilleumier et al., 2001).

To ensure that the main effect of trustworthiness did not arise from a highly significant activation in just one of the simple effects (i.e. that activation was task-independent) a mask was created from random effects SPMs for the simple effect of untrustworthiness under both tasks (each thresholded at $p<0.05$ uncorrected). This was used to mask the main effect of trustworthiness. Activations surviving this masking procedure reflect responses during both implicit and explicit judgements.

For the purposes of illustration, a second model was constructed by dividing the events for each subject into three groups by rank score for individual stimuli (i.e. the least trustworthy third of faces as one event type, the median third as a second and the most trustworthy third as a third event type). This model is used in Figures 6.3-6.5 to demonstrate the direction of BOLD signal change with respect to trustworthiness score. Note that statistical inferences are drawn solely from the parametric model described above.

The mean ratings of facial emotion derived from a second set of 16 age-matched subjects (see '*Emotion ratings for stimuli*' above) were used to construct another model for the data. In this model, subject-specific ratings for trustworthiness were entered as

parametric covariates, as before. Additionally, mean ratings for each of the four emotions (anger, fear, happiness and sadness) were entered as nuisance parametric covariates. The parameter estimates for trustworthiness are therefore rendered independent of the effects of the four facial expressions and variance better explained by the effects of a given facial expression will be attributed to the regressor modelling that facial expression. Contrast images for trustworthiness derived from this model were then entered into a random effects analysis.

Results

Behavioural

Post-scanning, subjects labelled on average more than half of the 120 faces as having 'neutral' emotional expressions (mean n=65). Labelled emotional expression interacted significantly with trustworthiness score across the group of subjects (Kruskal-Wallis test, $p<0.001$). Mann-Whitney U tests revealed that the trustworthiness scores assigned to 'disgusted, 'fearful' and 'surprised' faces did not differ significantly from 'neutral' faces ($p>0.05$ in all cases). 'Happy' faces (mean rating=4.0) were rated as significantly more trustworthy than 'neutral' faces (mean rating=3.9), and 'angry' (mean rating=2.7) and 'sad' (mean rating=3.6) faces as significantly less trustworthy ($p<0.01$ in all cases).

Means and standard deviations for the scanned cohort of subjects' ratings of trustworthiness are shown in Figure 6.1. Correlations between mean scores for facial emotion (from the second cohort of subjects – see 'Emotion ratings for stimuli' in Methods) and mean scores of trustworthiness of stimuli (from the scanned cohort) are

examined in Figure 6.2. Consistent with the data from the scanned cohort of subjects, significant correlations are shown between mean trustworthiness score and mean score for each of anger, happiness and sadness ($p<0.01$ two-tailed).

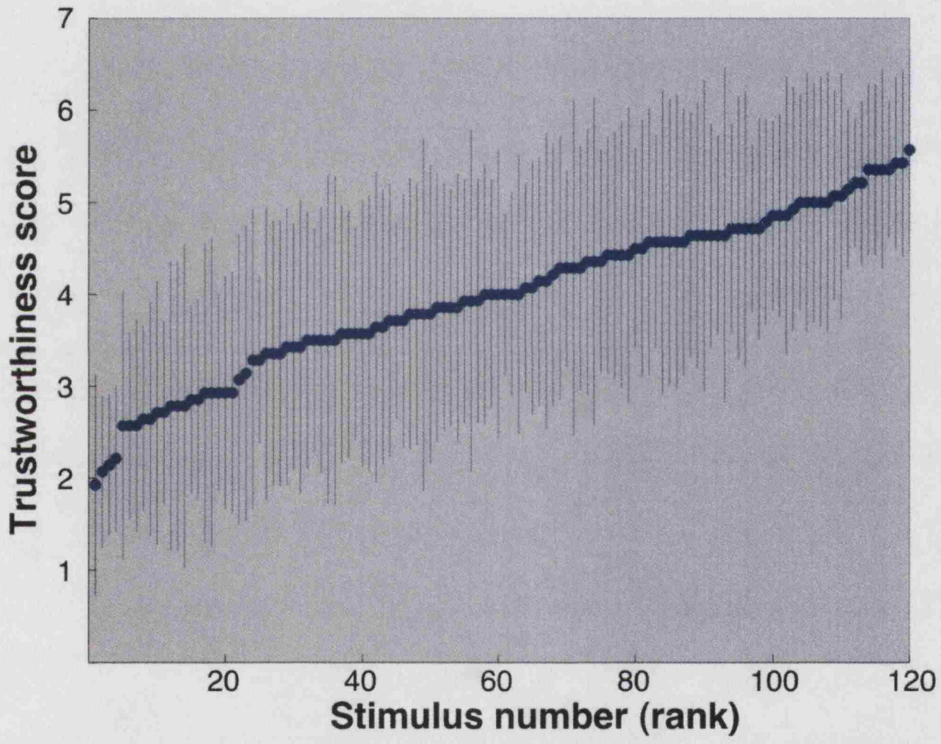

Figure 6.1: *Trustworthiness ratings for stimuli*

Means (diamonds) and standard deviations (error bars) of trustworthiness scores of stimuli rank-ordered by mean trustworthiness score across subjects

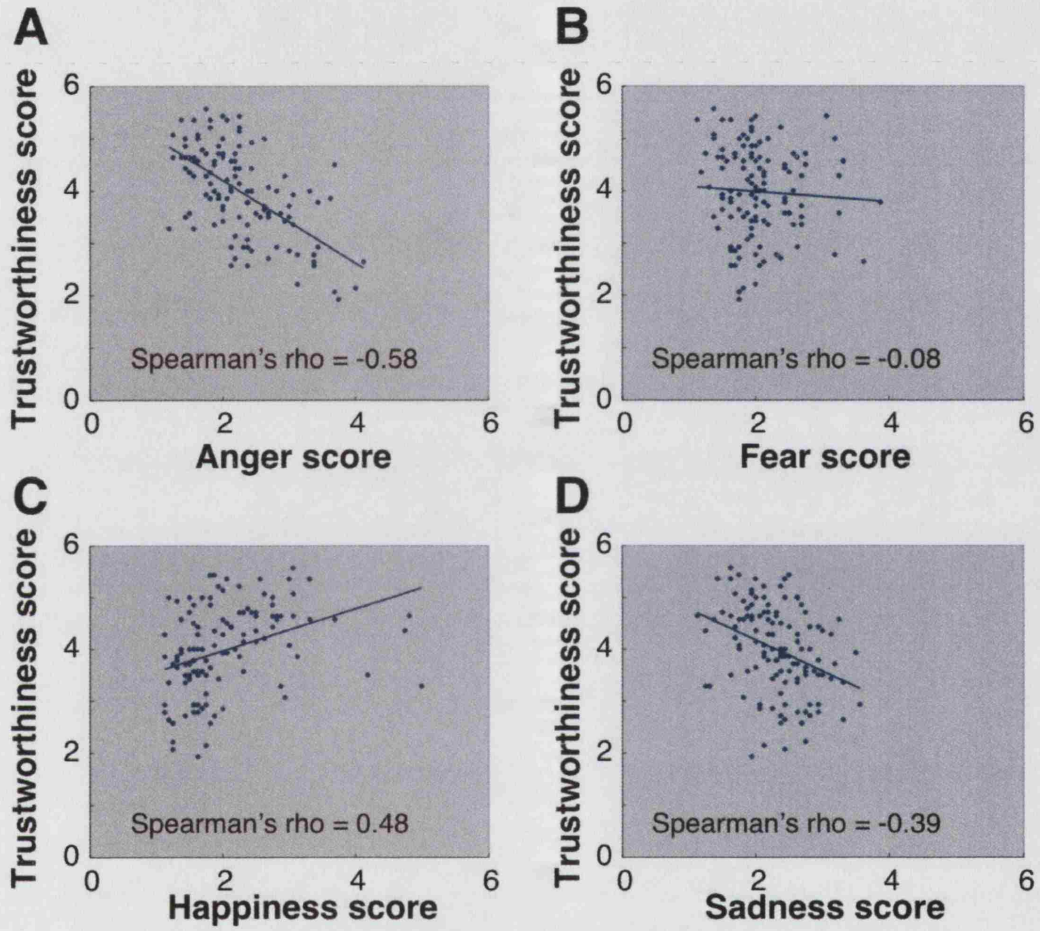

Figure 6.2: *Correlations between trustworthiness and emotion ratings*

Scatter plots of mean emotion scores (from second cohort of sixteen subjects) and mean trustworthiness scores (from cohort of subjects who underwent fMRI study) for stimuli: (a) anger, (b) fear, (c) happiness, (d) sadness. Lines of best fit are derived by linear regression. Both rating scales ranged from 1 (low degree of emotion or highly untrustworthy) to 7 (highly emotional or highly trustworthy).

Neuroimaging

Main effect of task

A significant activation in the explicit compared to implicit task, independent of trustworthiness was found in the right posterior superior temporal sulcus (x,y,z=56,-44,4; Z=4.27; $p<0.05$ SVC; Figure 6.3). Additionally, primary visual cortex was activated in this contrast ($p<0.001$ uncorrected). It has previously been demonstrated that attentional and emotional manipulations alter neural responses in early visual cortex (Lane et al., 1999) and similar processes engendered by the explicit task might account for this latter activation.

Main effect of trustworthiness

As predicted, significant bilateral amygdala activation was evident in the contrast of untrustworthy to trustworthy faces (right amygdala: x,y,z=-18,0,-24; Z=4.29; left amygdala: x,y,z=-16,-4.-20; Z=3.92; both $p<0.05$ SVC; Figure 6.4a). This examination of parametric data based upon each subject's ratings of faces indicates that the more untrustworthy the face the greater the BOLD response in the amygdala (Figure 6.4b).

Further areas showing increased response to untrustworthy faces included left superior temporal sulcus (x,y,z=-50,-58,10; Z=4.15) and a region of the right superior middle insula (x,y,z=42,-4,12; Z=3.48; Figure 6.4). Additionally, bilateral activation in the fusiform gyrus was evident in this contrast (right fusiform: x,y,z= 44,-46,-24, Z=3.58; left fusiform: x,y,z=-48,-48,-24; Z=3.60; both $p<0.05$ SVC; Figure 6.5).

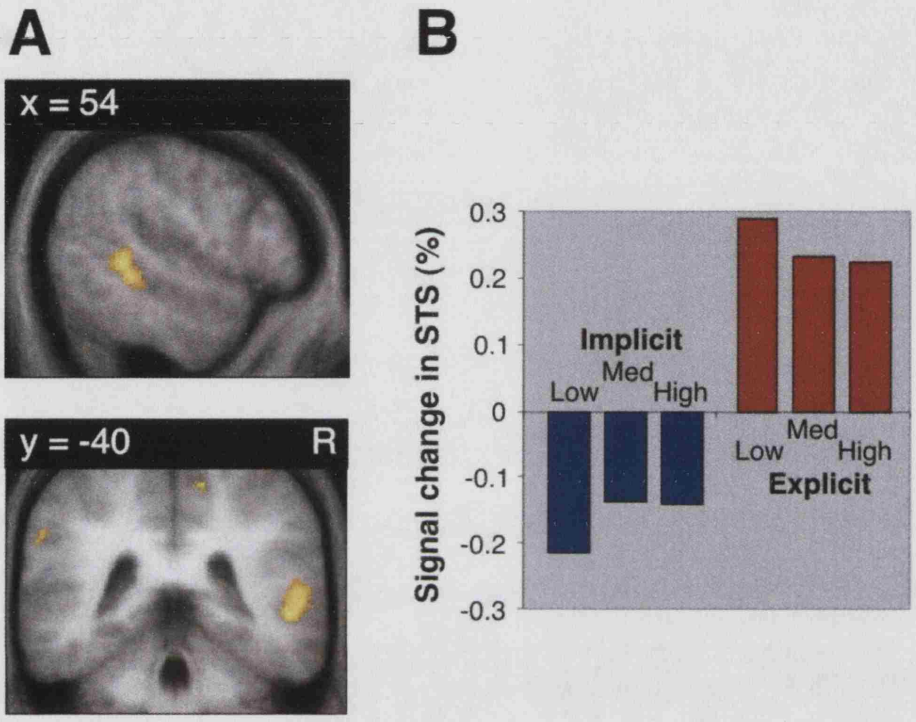

Figure 6.3: *Main effect of explicit social judgements in superior temporal sulcus*

a. Random effects SPM overlaid on averaged normalised structural scan from the group of subjects showing activation in right superior temporal sulcus (x,y,z = 56,-44,4; Z=4.27; $p<0.05$ SVC) when making judgements about trustworthiness compared to age. (Statistical threshold for display $p<0.001$ uncorrected; extent threshold of 5 voxels.)

b. BOLD signal measure categorised by task and trustworthiness of faces. "Low", "med" and "high" refer to the least trustworthy third, median third and most trustworthy third of faces calculated in the second model described in "Methods". y-axis represents mean (across subjects) percentage signal change relative to whole brain mean over scanning session for each event type. There is no clear pattern of response to the faces according to trustworthiness. Note that statistical inference is drawn only from the parametric model described in "Methods", and not from the illustrative model in the figure.

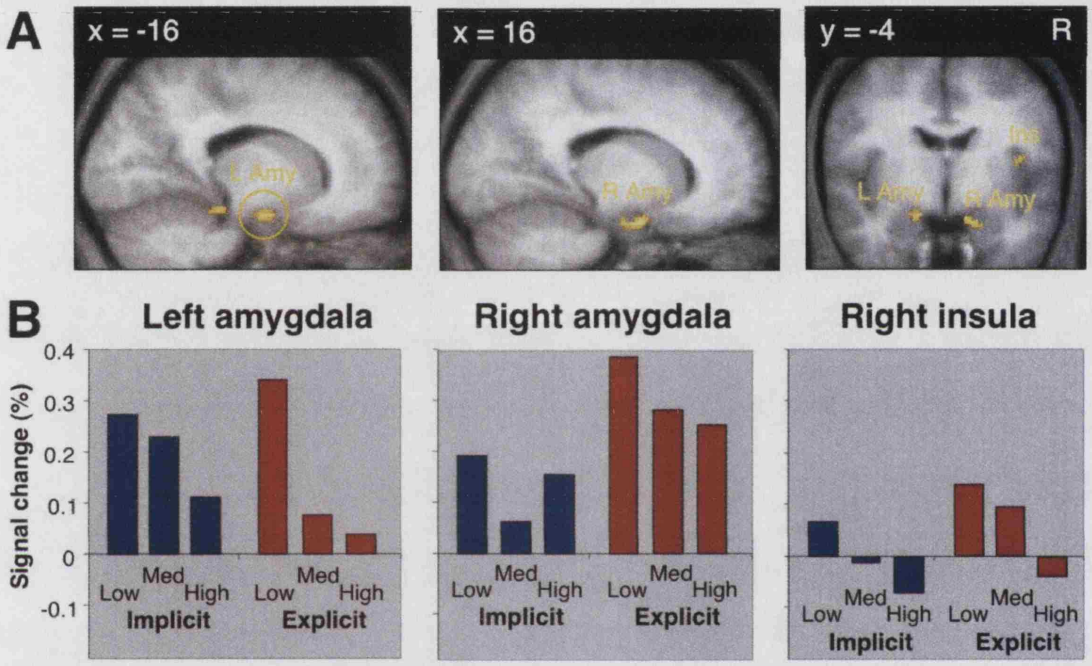

Figure 6.4: *Main effect of trustworthiness in amygdala and insula*

a. Significant increases in BOLD signal to untrustworthy faces in the right and left amygdala ("Amy") and right insula ("Ins"). (Right amygdala: x,y,z=-18,0,-24; Z=4.29; p<0.05 SVC; left amygdala: x,y,z=-16,-4.-20; Z=3.92; p<0.05 SVC; right insula (42,-4,12; Z=3.48; p<0.001 uncorrected). Display as in Figure 6.3a.

b. Responses to faces as a function of degree of individually rated trustworthiness illustrated for left amygdala, right amygdala and right insula. Note greater responses to less trustworthy faces across all these regions. y-axis as in Figure 6.3b.

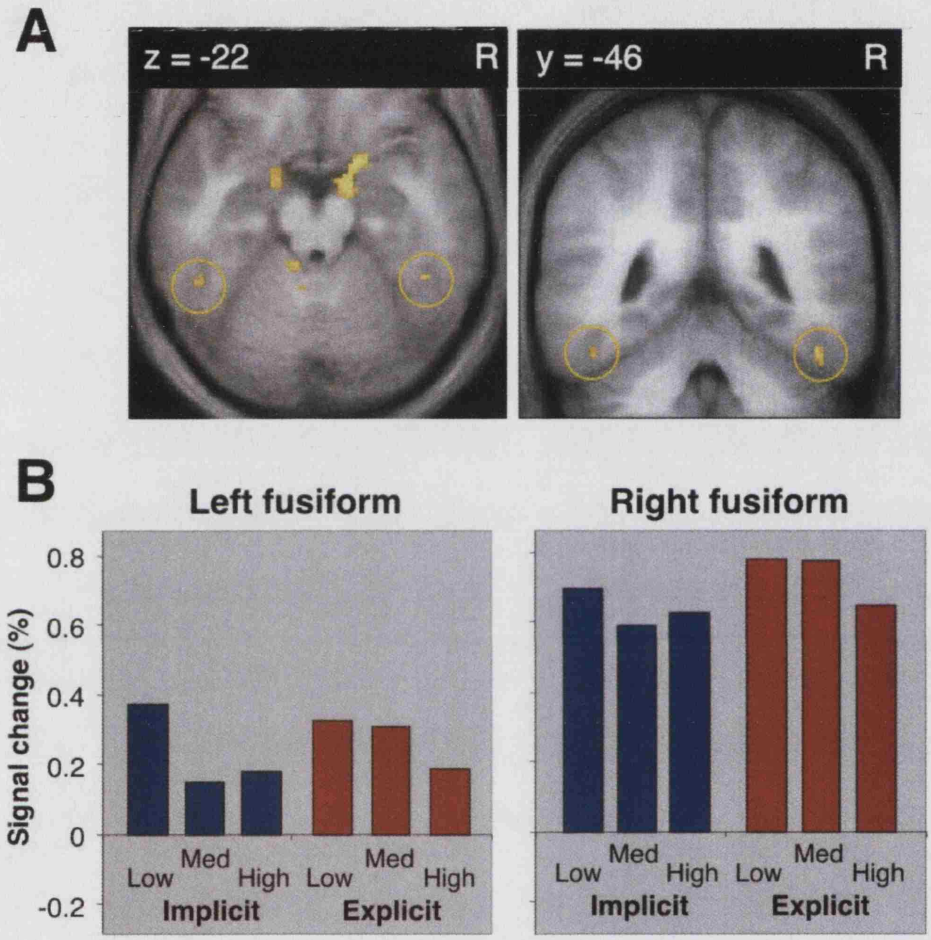

Figure 6.5: *Main effect of trustworthiness in fusiform gyrus*

a. Significant increases in BOLD signal to untrustworthy faces in the fusiform gyrus bilaterally (right fusiform: x,y,z=44,-46,-22; Z=3.58; $p<0.05$ SVC; left fusiform: x,y,z=-48,-48,-24; Z=3.60; $p<0.05$ SVC). Display as in Figure 6.3a.

b. Responses to faces as a function of degree of individually rated trustworthiness illustrated for left and right fusiform cortex. y-axis as in Figure 6.3b.

To ensure that the main effect of untrustworthiness was not driven by a highly significant activation in just one of the tasks alone, a masked conjunction of simple effects of trustworthiness under implicit and explicit task conditions was carried out. This confirmed that bilateral amygdala, fusiform gyrus and right insula showed significant responses to untrustworthy faces independent of task. Notably, left STS activation was not seen in this contrast and a post-hoc test revealed that the effects in this region were driven principally by trustworthiness judgements under the explicit task.

Interaction between task and trustworthiness

This contrast revealed an area in the lateral orbitofrontal cortex (OFC; x,y,z= -28,42,10; Z=3.73, $p<0.001$ uncorrected) responsive to untrustworthy faces in the implicit task and to trustworthy faces in the explicit task. However, this activation failed to survive correction for multiple comparisons across the entire volume of orbitofrontal cortex. No other regions of interest were revealed in this contrast, nor in the reverse interaction term.

Effects of trustworthiness independent of emotion

A subsidiary random effects analysis was performed across the 14 subjects using a model that modelled out effects from facial expression of basic emotions in the stimulus set. Even under these stringent criteria, right amygdala activation was still evident

(peak at x,y,z= 22,2,-18; Z=4.06; p<0.05 SVC; Figure 6.6). This activation overlapped with that reported in the primary model. At lower thresholds (p<0.005 uncorrected) there was additional activation in left amygdala.

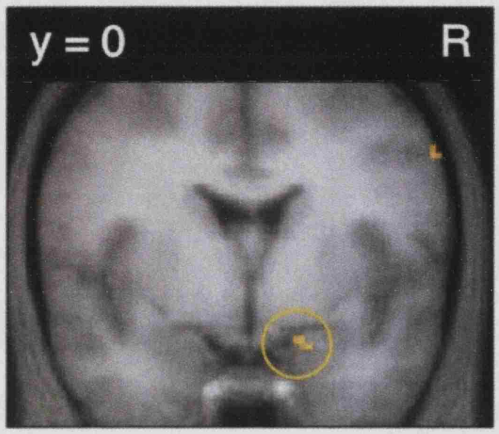

Figure 6.6: *Main effect of trustworthiness in amygdala independent of facial emotion*
Significant increases in BOLD signal to untrustworthy faces in right amygdala even when scores for four basic facial emotions are additionally used as parametric covariates in the analysis. This activation is significant at p<0.05 corrected for multiple comparisons across the volume of bilateral amygdala. Activation peak at x,y,z=18,2, -22; z=4.06, but overlaps with right amygdala activation focus shown in Figure 6.3a. At lower threshold of p<0.005 uncorrected, activation is evident in left amygdala.

Discussion

The question I addressed in this experiment was whether the dimension of trustworthiness in faces and the process of making social judgements are associated with distinct patterns of brain activation. The finding of activation of amygdala, orbitofrontal cortex and STS is highly consistent with the model of Brothers (1990). The current results also extend upon previous lesion data (Adolphs et al., 1998), by showing amygdala activity to untrustworthy faces regardless of whether subjects were

explicitly making trustworthiness judgements. This finding echoes earlier studies of obligatory threat-related processing in the amygdala (Morris et al., 1998b; Whalen et al., 1998; Morris et al., 2001b; Vuilleumier et al., 2001; see also Chapter 4). In contradistinction to many imaging studies of facial emotion (e.g. Morris et al., 1996) the amygdala activation to untrustworthy faces was bilateral, supporting neuropsychological evidence that patients with unilateral amygdala lesions can successfully make trustworthiness judgements (Adolphs et al., 1998).

In addition to amygdala, right insula was also activated by faces that subjects considered untrustworthy regardless of task. The insula has been shown to be activated in a wide variety of functional imaging studies of emotion (e.g. Buchel et al., 1998b, 1999; Casey, 1999; Critchley et al., 2001a, 2002). One suggested role for insula is the mapping of autonomic changes as they affect the body where such mappings form the basis of 'gut feelings' about emotive stimuli (Damasio, 1999; Critchley et al., 2001b). Thus, a possible explanation for the observed insula activation is that a consequence of amygdala activation is the generation of autonomically mediated changes in bodily states which are then re-mapped to the insula.

Differential activation in face-responsive regions of the fusiform gyrus was observed in relation to the dimension of trustworthiness in face stimuli. Previous imaging studies have demonstrated increased activity in modality-specific cortical areas to stimuli with emotional content relative to non-emotional stimuli (e.g. Breiter et al., 1996; Isenberg et al., 1999; Strange et al., 2000; Morris et al., 2001a; Vuilleumier et al., 2001; see also Chapters 4 and 5). Enhanced fusiform activation in response to emotive face stimuli has been attributed to modulatory influences from amygdala (Morris et al., 1998a)

possibly mediated by anatomical back-projections (Amaral et al., 1992). Indeed, a human lesion study highlights a possible role for amygdala in enhancing perceptual processing of threat stimuli (Anderson and Phelps, 2001). Such processes might extend to faces representing potential threat at the social level and a neural consequence is enhanced fusiform activation.

Right STS showed task-related activation in the explicit judgement condition but no differential activity according to the dimension of trustworthiness. In other words, right STS was activated when subjects made explicit judgements about trustworthiness. In this regard the STS showed activity when subjects were required to make inferences concerning the likely intentionality of others. This region has been implicated in functional imaging studies on biological motion (Bonda et al., 1996), and biological-like motion (Castelli et al., 2000). More critically, studies have found activity in posterior STS and adjacent regions at the temporo-parietal junction when subjects make theory of mind inferences (Fletcher et al., 1995; Brunet et al., 2000; Gallagher et al., 2000). It has been suggested that this region is involved in intention detection (Frith and Frith, 1999), rather than biological motion processing per se. Intention detection might form a critical component in determining whether or not to trust an individual, and activity in this region could be explained by such a process.

Evidence from human patients with discrete lesions of orbitofrontal cortex (OFC) indicate that this region is critical for complex social judgement (Eslinger and Damasio, 1985; Damasio, 1994). Unlike in amygdala, I found that activation in this region was task-dependent. When subjects made explicit judgements of trustworthiness this region showed an enhanced response to faces deemed trustworthy. By contrast, when judging

age, this region showed greater responses to untrustworthy individuals. Notably, other studies have reported similar task-dependent responses in lateral orbitofrontal cortex. For example, differential responses in a region of lateral OFC are evident when a stimulus is subject to preference or recognition judgement (Elliott and Dolan, 1998). A dissociation between implicit/automatic social judgement and explicit (laboratory-tested) social judgement has also been reported in a patient with orbitofrontal cortex damage (Eslinger and Damasio, 1985). This patient remained able to evaluate social situations under explicit task instructions but was impaired in day-to-day ("automatic" - Eslinger and Damasio, 1985) social judgements. Note that activation in this region did not survive correction for multiple comparisons and effects in this region are emphasised on the basis of its known involvement in social judgements (Eslinger and Damasio, 1985; Damasio, 1994).

Several regions of interest in this study (amygdala, OFC and insula) have previously been shown to activate in processing specific facial expressions, or facial expressions generally (Chapter 4). Facial expressions of fear consistently activate the amygdala (Breiter et al., 1996; Morris et al., 1996; Phillips et al., 1997; Whalen et al., 1998; Vuilleumier et al., 2001), while facial expressions of disgust activate anterior insula (Phillips et al., 1997, 1998). Additionally, correlations between the trustworthiness scores and scores for facial emotions attributed to the stimulus set were evident (Figure 6.2). Consequently, one possibility is that differential patterns of activation seen in this study reflect influences from one or more emotional expressions alone. This possibility was assessed by analysing the fMRI data with additional nuisance covariates pertaining to the degree of emotional expression of each of four basic emotions (anger, fear, happiness, and sadness). A significant right amygdala response to untrustworthy faces

persisted, despite the fact that variance attributed to facial emotion was now accounted for in this secondary analysis. These results suggest that facial expressions of emotion provide a constituent element in making trustworthiness judgements but that amygdala responses also showed independence of these effects. Notably, the deficit of patients with bilateral amygdala lesions in making social judgements is in the context of maintained ability to utilise information about emotional expression of the face stimuli (Adolphs et al., 1998).

Summary

The study reported in this chapter has found evidence that extends the neural substrates of face perception characterised in earlier chapters to processing the complex social characteristic of trustworthiness from faces. The data extend the model of Brothers (1990), highlighting a dissociation between automatic and intentional engagement within this proposed circuitry of this model. Thus, social judgements, in relation to faces, reflect a combination of brain responses that are, in the case of the amygdala, stimulus driven and in the case of STS driven by processes relating to inferences concerning the intentionality of others.

Chapter 7: Neural responses to attractiveness in faces

Introduction

The previous chapters demonstrated that systems in the brain automatically decode facial features such as expressions of emotion and trustworthiness. In this chapter, I explore the nature of the brain's response to attractive and unattractive faces, testing a number of different hypotheses. Specifically, I examine the task-dependence of the response of regions identified in previous studies reported in the literature as processing facial attractiveness. Additionally, I ask why these previous studies addressing facial attractiveness processing have not demonstrated responses in the amygdala, and test the hypothesis that this is because these early studies have been limited by assumptions of linearity of response in the amygdala across "attractiveness space". As attractiveness is a characteristic judged differently by males and females, a large study cohort was used and cross-gender differences tested.

Human facial attractiveness conveys significant biological advantages as expressed in mating success, earning potential (Frieze et al., 1991) and longevity (Henderson and Anglin, 2003). It can be conjectured that facial attractiveness is an important variable in mate choice (Thornhill and Gangestad, 1999; Fink and Penton-Voak, 2002) and that evolved brain systems show sensitivity to this aspect of the sensory environment.

The idea that the human brain possesses regions responsive to attractiveness is supported by data from brain imaging studies. Aharon et al (2001) showed that attractive female faces activate reward regions in males more than attractive males or

unattractive faces of either gender. O'Doherty et al (2003b) showed that dissociable regions of prefrontal cortex responded to attractive and unattractive faces; specifically showing that medial prefrontal regions, including medial orbitofrontal cortex, responded to attractive faces and lateral regions respond more to unattractive faces. The theoretical approach in both these studies was to treat viewing of attractive faces as akin to reward, an approach vindicated by behavioural data (Aharon et al., 2001) demonstrating that male subjects work to observe attractive female faces, but not unattractive females or any male face. In addition, behavioural evidence demonstrates that an attractive female face will lead males to discount higher future rewards against smaller immediate rewards (Wilson and Daly, 2004), consistent with this construal of attractive faces as rewards.

In brain imaging studies (Aharon et al., 2001; O'Doherty et al., 2003b) the use of indirect tasks to disguise the key experimental variable of facial attractiveness suggests that observed responses had a strong obligatory component. An aim in the experiment described in this chapter was to assess the degree of independence in the evoked response with respect to task. In other experiments described in this thesis, this approach has indicated that dissociable brain regions are invoked when processing facial characteristics such as expression and trustworthiness compared to performing tasks wherein the variable of interest is orthogonal to the emotive facial component (Chapters 4 and 6). However, such an approach is rarely adopted in the study of attractiveness or reward more generally (but see Nakamura et al., 1998; Zatorre et al., 2000; Royet et al., 2001).

I also wished to address another limitation of previous studies, namely their use of predefined categorisation of faces as attractive or unattractive and an assumption of linearity in the evoked brain response. Although in O'Doherty et al (2003b), a categorical analysis was supplemented by a parametric exploration of linear effects, the dataset was not optimised to examine potential non-linear relationships between attractiveness and neural activity. Thus, in the present study I investigated whether evoked responses expressed as a quadratic form, showing enhanced responses to both attractive and unattractive faces, would be evident. Despite the fact that amygdala is frequently activated in processing aversive or appetitive stimuli (Baxter and Murray, 2002; Sander et al., 2003; Zald, 2003), previous neuroimaging studies concerning attractiveness have failed to demonstrate activation in this region. I hypothesised that previous approaches to the study of the role of the amygdala in sensing attractiveness have been limited by an assumption that it responds linearly across "valence space" and that instead amygdala might index relative value such that there is maximal activity at extremes of valence. This is predicted based upon the evidence in Chapter 4 that amygdala responds to a variety of different facial expressions of emotion, and was tested here by examination of nonlinear relationships between BOLD and subjective attractiveness.

I used a factorial fMRI experimental design with a parametric factor of attractiveness, and categorical factors of task, stimulus (face) gender and subject gender. In addition to a replication of previous findings (Aharon et al., 2001; O'Doherty et al., 2003b) I hypothesised that regions responsive to attractive faces would show an enhanced response under the direct task condition, and that amygdala would demonstrate responses non-linearly related to attractiveness. Aside from amygdala, OFC and

striatum, the key regions of interest in this report are the core components of the distributed system for face perception (Haxby et al., 2000), namely fusiform cortex and superior temporal sulcus (STS).

Methods

Stimuli

I supplemented the stimuli from O'Doherty et al (2003b) with additional faces from similar sources. The additional stimuli were picked primarily to be of average attractiveness, as the existing stimuli in the set were generally considered to represent extremes of the attractiveness spectrum. Stimuli, which were colour image files, were cropped similarly to the originals with little hair visible and adjusted to have equal mean luminance. Faces had direct eye-gaze and head direction and were approximately coregistered for eye-position. As in the previous study, images showed expressions between neutral and mild smiles. Seventy-two images were used (36 females).

Subjects

fMRI data was obtained from 28 subjects (13 females). Data from two subjects (both males) were excluded after a debriefing questionnaire revealed non-heterosexual preferences (self-rated sexuality <5 on scale where 1=exclusively homosexual and 7=exclusively heterosexual). The age range of the remaining subjects was 18-35 (mean = 25.5), with no significant difference between the ages of males and females ($p=0.12$).

Experimental paradigm

Subjects performed one of two tasks on the face stimuli, with task alternating in a blocked fashion. In the *attractiveness* task subjects made a judgement of attractiveness with one of three buttons ("highly attractive", "medium", "low attractiveness"). In the *age* task subjects judged the age of the face as being young, medium or old, with three buttons to represent the response. Blocks of each task were preceded by an instruction and a summary instruction remained onscreen during each block. Stimuli were presented for 1000ms and inter-stimulus interval (ISI) was 1900ms. Blocks consisted of nine stimuli giving an overall block length of 26s. In total, there were 16 blocks for each task and each stimulus was presented twice in the context of each task. The starting block was counterbalanced across subjects.

fMRI data

Data were collected on a 1.5T MRI scanner using gradient echo T2*-weighted echo-planar images, with blood oxygenation level dependent (BOLD) contrast. Volumes consisted of 36 slices angled at -30° to the horizontal. The effective repetition time (TR) was 3.2s, and 295 volumes were collected, with the first five subsequently discarded. A T1-weighted structural image was acquired for each subject for detailed anatomical information.

Debriefing

Participants undertook two debriefing tasks outside the scanner. They first rated all the faces on attractiveness, using a computerised visual analogue scale. Subsequently they rated the faces on happiness on a similar scale with the extremes labelled "happy" and "unhappy", and the mid-point indicated. Participants also provided information concerning their sexual orientation on an ordinal scale.

Data analysis

Imaging data were pre-processed and analysed in the standard manner described in the general methods section using SPM2. Preprocessing included realignment, slice time correction, normalisation and smoothing. In the statistical analysis, the approach was standard, with the exception of the inclusion of parametric modulation to represent the attractiveness and happiness variables. Attractiveness was represented using four polynomial expansions of the rating for each face (see Buchel et al., 1998a), happiness with just a linear term and the interaction between happiness and attractiveness represented by the dot-multiplication of the two scores. Response times were modelled as a parametric covariate of no interest. Random effects analysis was used for inference.

In all models reported (with the exception of conjunction models), males and females were treated as separate groups in an ANOVA model and equal variance was not assumed. For conjunctions, males and females were treated as a single group and non-sphericity correction was applied for the non-independence of the contrasts of parameter

estimates from this repeated measure design. $p<0.05$ corrected for multiple comparisons across the brain was considered the threshold for statistical significance. Exceptions were made in *a priori* regions of interest, where the more liberal threshold of $p<0.001$ uncorrected was used with a correction for multiple comparisons over a small region of interest (Worsley et al., 1996). Regions of interest were medial and lateral orbitofrontal cortex, medial prefrontal cortex and anterior insula, based upon a previous study (O'Doherty et al., 2003b). In addition regions considered to comprise the core system for human face perception (Haxby et al., 2000), namely fusiform cortex and superior temporal sulcus were considered a priori regions of interest. Finally, the amygdala was a prime region of interest, given the prediction of non-linear responses in this region. Medial OFC and inferior medial PFC were mapped using the mean anatomical images and delineated to include gyrus rectus, frontomarginal gyrus and medial orbital gyrus inferiorly and supraorbital gyrus medially and superiorly, with a total mask volume of $20cm^3$. The amygdala was delineated bilaterally on the mean anatomical images, and the total mask volume was $8cm^3$.

Behavioural data during scanning

Ratings of attractiveness during the attractiveness task (for comparison to debriefing ratings) and response times for both conditions were obtained in every subject except one male owing to technical failure. Responses faster than 200ms or slower than 2 standard deviations above the within-subject mean were excluded from further analysis. Pupillometric data were available in 16 out of the subjects (six females). Eye blinks were removed by interpolation between points where pupil diameter was below 80% of the mean signal (excluding zeros) across the epoch. These behavioural data were

analysed in a two-stage model similar to the fMRI analysis. Thus, in a first stage procedure the data were modelled with parametric polynomial expansions pertaining to attractiveness and linear trends for happiness and the interaction between happiness and attractiveness. A second stage examined for consistency in these effects across subjects using ANOVAs. Three measures of pupil data were taken, namely the mean, minimum and maximum of response during a 2033ms window after presentation of the face stimulus. In addition, trial-by-trial measures of response times (and pupillometry in a supplementary analysis) were used as parametric covariates of no interest in fMRI data analysis to ensure BOLD effects were not a result of behavioural "confounds".

Results

Behavioural

Ratings of attractiveness during the scanning session correlated well with ratings from debriefing in 24 subjects (out of 25 for whom data were available from the scanning session). Across the group the mean slope of this regression was positive and significantly different from zero (t_{24}=17.0, $p<0.001$) indicating strong within-subject agreement for ratings obtained inside and outside the scanner.

The stimuli represented a broad range of attractiveness (Figure 7.1). Two measures suggest that ratings of attractiveness and happiness were significantly (positively) correlated across the group of subjects (Figure 7.2). First, there was a marginally significant correlation between the mean ratings of happiness and attractiveness across

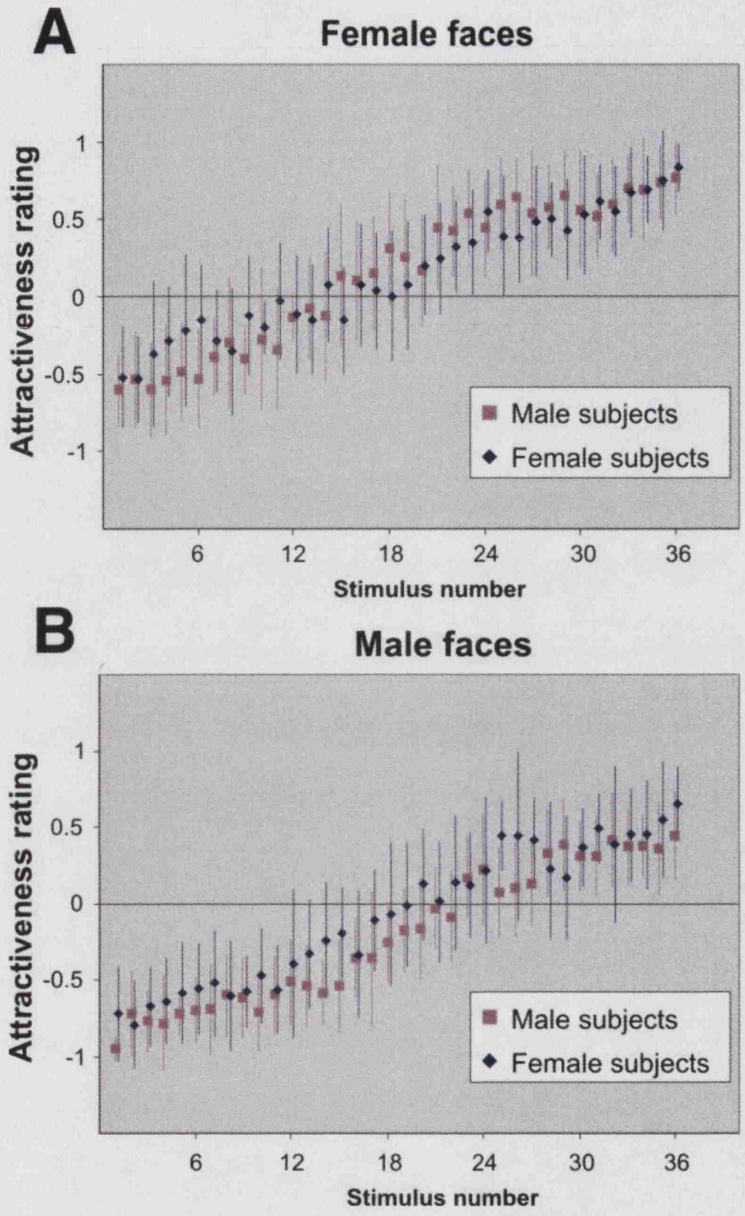

Figure 7.1: *Behavioural ratings of facial attractiveness*

Faces spanned a range of attractiveness ratings for both male and female subjects. Here, stimuli are rank-ordered by mean score across subject gender for female (a) and male (b) faces, and mean ratings for male and female subjects are given separately. There was good inter-subject and inter-gender agreement. Error bars are standard deviation.

the 72 stimuli (p=0.050). Second, the across-subjects average slope of within-subject regression of happiness on attractiveness ratings was significantly above zero (t_{25}=6.05; p<0.001). There was no significant difference between males and females with regard to the magnitude of this correlation (p=0.80). This correlation between attractiveness and happiness rendered this experiment relatively insensitive to the interaction between perceived smile and attractiveness previously described (O'Doherty et al., 2003b). Since the primary interest was attractiveness, happiness ratings entered into the design matrix for fMRI data were orthogonalised with respect to attractiveness.

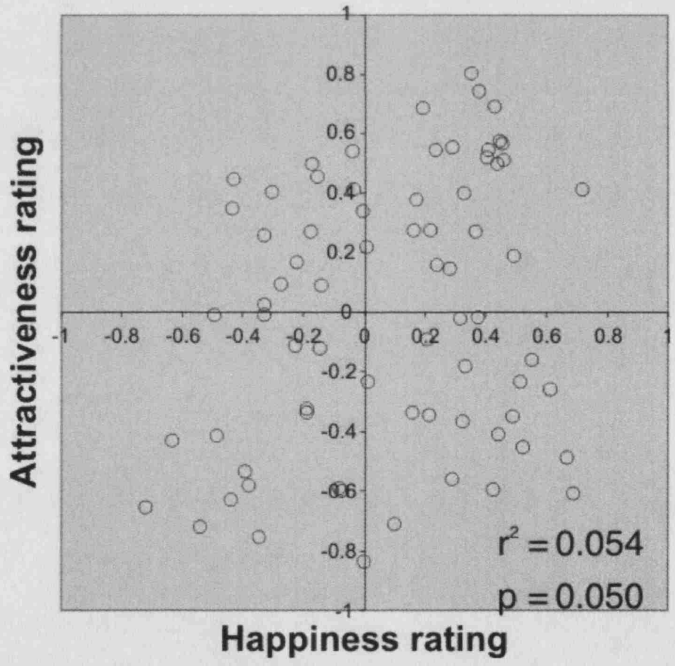

Figure 7.2: *Correlation between ratings of attractiveness and happiness*

There was some positive correlation between ratings of attractiveness and ratings of happiness, both within (see text) and between subjects.

Analysis of response time (RT) data during the fMRI experiment revealed a highly significant main effect of task ($F_{1,23}$=43.16; p<0.001) and a significant task by stimulus (face) gender interaction ($F_{1,23}$=7.05; p=0.014), driven by faster responses in the attractiveness task (mean RT in attractiveness task = 1043ms; mean RT in age task = 1139ms) with an exaggerated difference for female relative to male faces in the attractiveness task alone (mean difference between male and female faces in age task = 16ms; t_{24}=1.50; p=0.15; mean difference between male and female faces in attractiveness task = 44ms; t_{24}=6.49; p<0.001). Given that RTs were modelled in the analysis of fMRI data, linear effects of this confounding variable cannot explain the imaging results reported below.

No significant main effect was seen in analysis of pupillometry data, indicating that there was no overall simple (first or second order) relationship between pupil dilation and attractiveness rating. Interestingly, there were interactions between attractiveness and subject gender for two of the three measures of pupil response tested (mean [p=0.02] and maximum [p=0.007]). In both cases the interaction was driven by a significant correlation between attractiveness rating and pupil response in the male subjects that was not present in females (mean: males one-sample t-test, t_9=-3.45, p=0.007; females one-sample t-test, t_5=0.92, p=0.40; maximum: males one-sample t-test, t_9=-5.29, p<0.001; females one-sample t-test, t_5=1.03, p=0.35). This indicates that male subjects showed arousal profiles linearly sensitive to attractiveness (showing greater pupil dilation for less attractive faces) whereas females did not.

Neuroimaging

Main effect of attractiveness

A significant positive effect of attractiveness (greater attractiveness associated with linear increases in BOLD) was seen in anterior cingulate (x,y,z = -3,36,0; Z = 4.89; $p<0.05$ corrected; Figure 7.3). However, this main effect was driven primarily by a subject gender by attractiveness interaction, as the effect of enhanced cingulate response to attractive faces was expressed solely in male subjects (x,y,z = -3,36,0; Z for interaction = 3.71; $p<0.001$ uncorrected two-tailed; Figure 7.3). Left mid-insula showed positive linear effects of attractiveness that bordered significance (x,y,z = -42,0,15; Z=4.71; p=0.063 corrected). Enhanced activations to attractiveness in *a priori* regions of interest included left posterior occipito-temporal cortex extending into lateral fusiform (x,y,z = -48,-66,-15; Z=4.27; $p<0.001$ uncorrected; fusiform peak: x,y,z = -39, -69,-12; Z=3.92; $p<0.001$ uncorrected), and orbitofrontal cortex (x,y,z = -21,54,-9; Z=3.97; $p<0.05$ small volume corrected (SVC); x,y,z = -15,48,-15; Z=3.11; $p<0.001$ uncorrected; Figure 7.6). This latter region showed significant effects in male and female subjects and across tasks, as evidenced by conjunction analyses.

No region showed significant effects for the linear effect of unattractiveness.

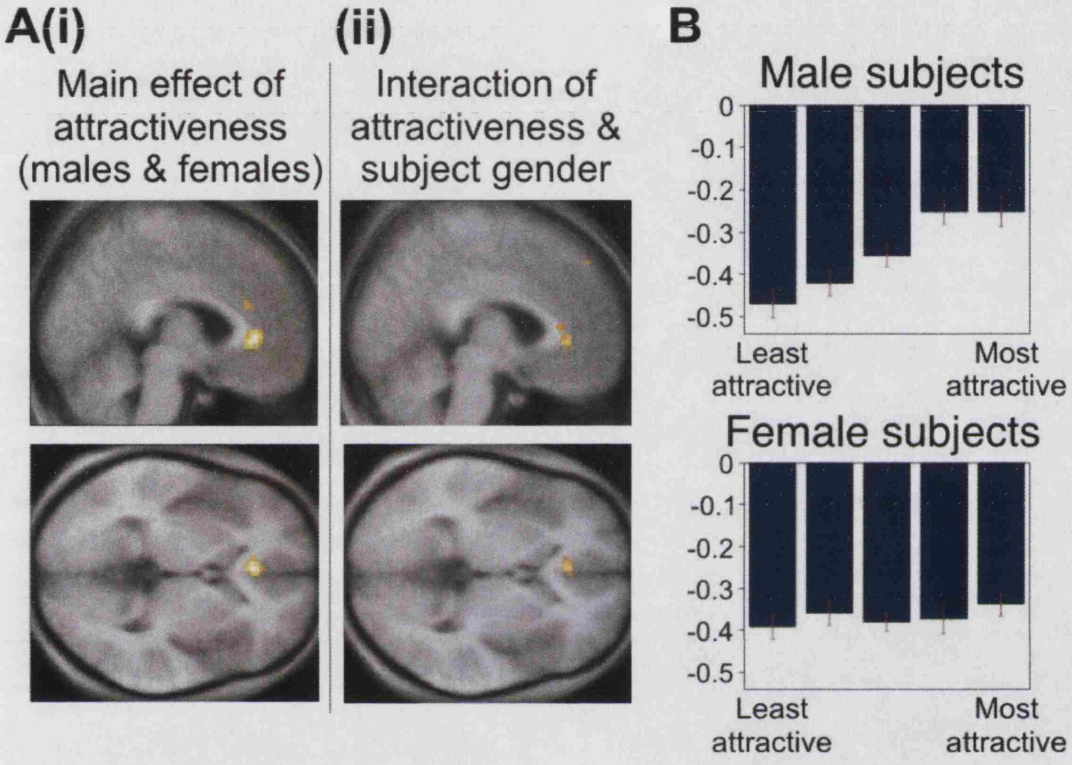

Figure 7.3: *Anterior cingulate shows a subject gender-by-attractiveness interaction*

a. (i)Although anterior cingulate shows a significant main effect of attractiveness, this is clarified by an attractiveness-by-subject gender interaction (ii). (i) and (ii) are Statistical Parametric Maps (SPMs) of the main effect of attractiveness (attractive > unattractive) and interaction between attractiveness and subject gender, respectively. SPMs are overlaid on mean T1-weighted image from the 26 subjects; $p<0.001$ is used as the threshold for display.

b. Only males show a clear pattern of responding more to attractive than unattractive faces. Data are derived from a subsidiary model in which stimuli were divided into five equally-sized groups for each subject based upon attractiveness scores and fMRI data fitted to this model. Parameter estimates of size of response are averaged for each cohort of faces. Note that there was no baseline condition nor any null event in the experimental design so responses can be interpreted only with respect to one another, not with respect to zero. (Units are % signal change.)

Main effect of task

Significant main effects of task were seen in a number of regions (Table 7.1) including bilateral superior temporal sulcus (STS) (attractiveness task > age task) along the A-P axis (Figure 7.4). A large region of superior medial prefrontal/paracingulate cortex showed a similar effect as did bilateral posterior orbitofrontal cortex (OFC)/anterior insula (Figure 7.5). The activation of left posterior OFC/insula extended posteriorly (at uncorrected thresholds) into amygdala (peak at x,y,z = -24,0,-24; Z = 4.11; $p<0.001$ uncorrected).

In the opposite contrast, significant effects resulting from greater responses during the age task were seen in right posterior (retrosplenial) cingulate cortex (x,y,z = 12,-51,18; Z=5.84; $p<0.05$ corrected) alone. At uncorrected thresholds a similar region was also seen on the left (x,y,z = -9,-54,12; Z = 4.45; $p<0.001$ uncorrected).

Task by attractiveness interaction

Inferior medial prefrontal cortex close to the region shown to respond to attractiveness during gender judgements by O'Doherty et al (2003b) showed an interaction between task and attractiveness (x,y,z = 0,60,-6; Z=3.68; $p<0.001$ uncorrected; Figure 7.6). This interaction took a cross-over form, with positive relationship between attractiveness and activation during the age judgement task ($p<0.001$, two-tailed t-test) and a negative correlation under the attractiveness judgement task ($p<0.028$, two-tailed t-test).

Region		Coordinates (mm)			Significance (Z)	Cluster size (voxels)
		x	y	z		
STS	Left	-60	-45	9	5.94	39
		-48	-42	3	4.95	
	Right	54	-48	27	5.20	7
		51	-21	-6	5.48	6
Posterior OFC	Left	-36	21	-24	5.66	34
(Inferior frontal gyrus)		-51	24	-6	5.09	
	Right	36	24	-21	5.13	3
		36	30	-18	4.85	1
Paracingulate cortex/		0	51	18	5.65	155
Medial prefrontal cortex		-6	60	15	5.49	
		3	54	30	5.18	
Superior frontal gyrus	Left	-9	27	60	5.59	9
		-18	54	39	5.15	9
		-18	45	42	5.09	
		-18	42	36	4.94	2
	Right	15	21	63	4.88	1
Anterior cingulate		3	36	9	5.45	9
Posterior cingulate		0	-18	39	5.15	10
	Left	-6	-48	36	5.00	3
Inf. temporal cortex	Right	54	0	-36	5.04	6
Occipital cortex	Right	6	-96	9	4.94	2
Cerebellum	Right	21	-84	-24	4.83	1

Table 7.1: *Brain regions showing significant differences in BOLD activation in a task dependent manner*

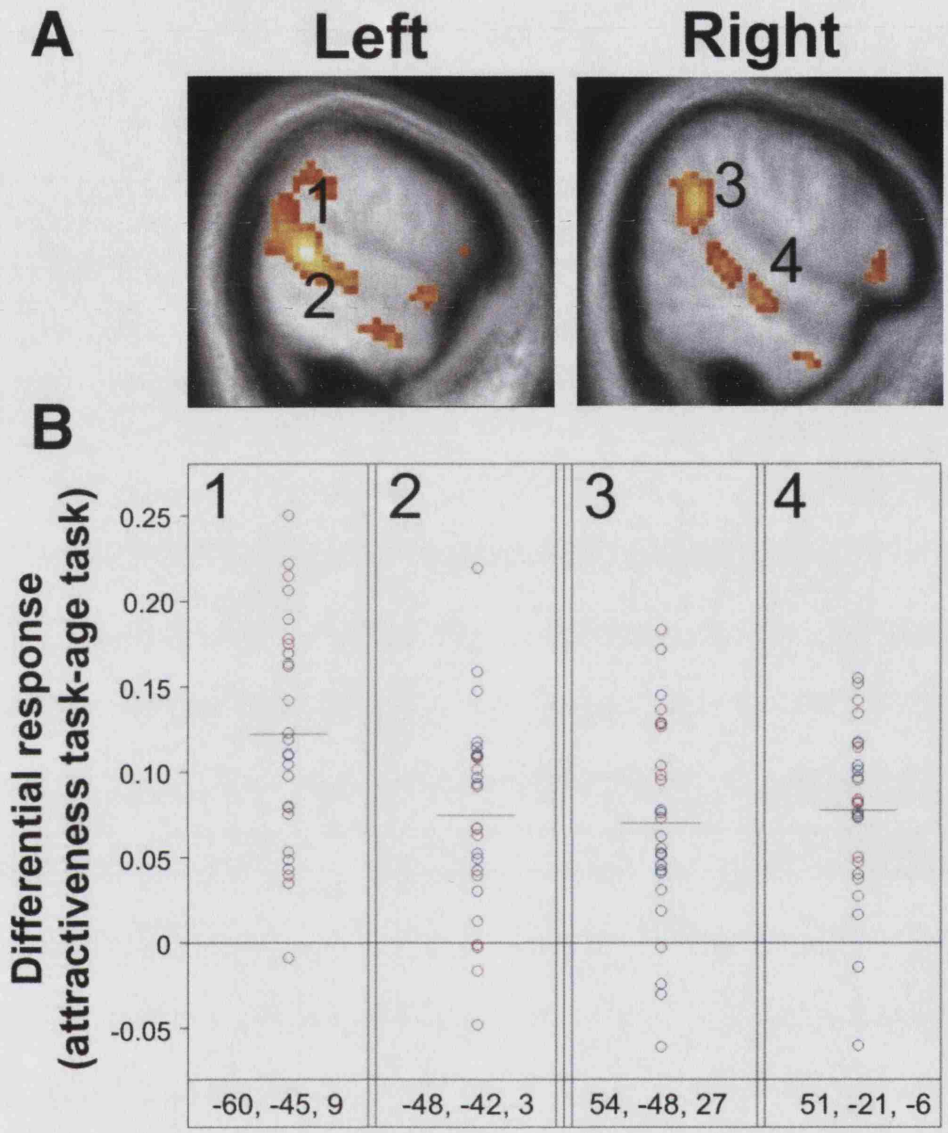

Figure 7.4: *STS shows enhanced responses during attractiveness judgements*

a. SPM showing main effects of task (attractiveness judgement > age judgement). Display as if Figure 7.3a. Four peaks within STS (labelled 1-4) show significant effects corrected for multiple comparisons across the whole brain.

b. Differential effects for each peak are shown in the plots: red circles represent male subjects and blue circles females. The horizontal line shows mean differential response between the two tasks. (Units are % signal change.)

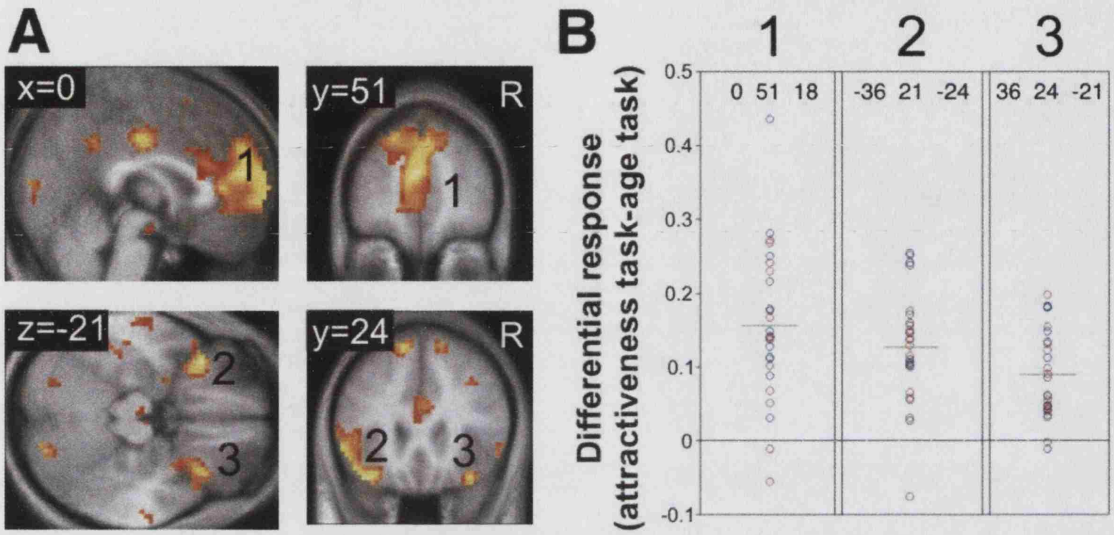

Figure 7.5: *Prefrontal regions show enhanced responses during attractiveness judgements*

a. SPM showing main effects of task (attractiveness judgement > age judgement). Three loci (labelled 1-3) in prefrontal cortex show responses significant corrected for multiple comparisons across the whole brain. These include medial prefrontal cortex (1), and bilateral orbitofrontal cortex (2, 3) extending posteriorly into insula and amygdala on the left. Display as in Figure 7.3a.

b. Differential effects for peaks 1-3. Red circles represent male subjects and blue circles females. The horizontal line shows mean differential response between the two tasks.

Although no regions showed significant interactions between subject gender, face gender and attractiveness, a region of inferior medial prefrontal cortex showed a significant four way interaction between attractiveness, face gender, task and subject gender (x,y,z = 3,57,-12; Z=3.38; p<0.001 uncorrected; Figure 7.6). This interaction took a cross-over form, and was driven by a greater linear increase in attractiveness responses to opposite sex faces for subjects of either gender in the age task compared to the attractiveness task.

Non-linear effects of attractiveness

Significant positive second order relationships between BOLD activation and attractiveness score were seen in amygdala (x,y,z = 27,0,-24; Z=3.52; p<0.05 SVC; Figure 7.7), right middle temporal gyrus adjacent to the lower bank of the STS (x,y,z = 60,-42,0; Z=4.39; p<0.001 uncorrected), and medial OFC (x,y,z = -3,30,-27; Z=3.60; p<0.001 uncorrected; Figure 7.6). These regions showed greater activation when subjects observed highly attractive or unattractive faces than faces of medium attractiveness. All of these activations were task-independent, as evidenced by significant results in a conjunction analysis across the factor of task. The response in the amygdala was independent of subject gender or face gender, as shown by conjunction analyses across these factors (x,y,z=27,0,-21; Z=3.91; p<0.05 SVC). Visible in Figure 7.7 is a trend to unattractive faces yielding greater activation than attractive faces, consistent with a marginal negative regression slope for the linear term at the peak voxel from the non-linear effect (Z = 1.80, p<0.05 uncorrected). Interestingly, the non-linearity appears less prominent in male subjects observing male

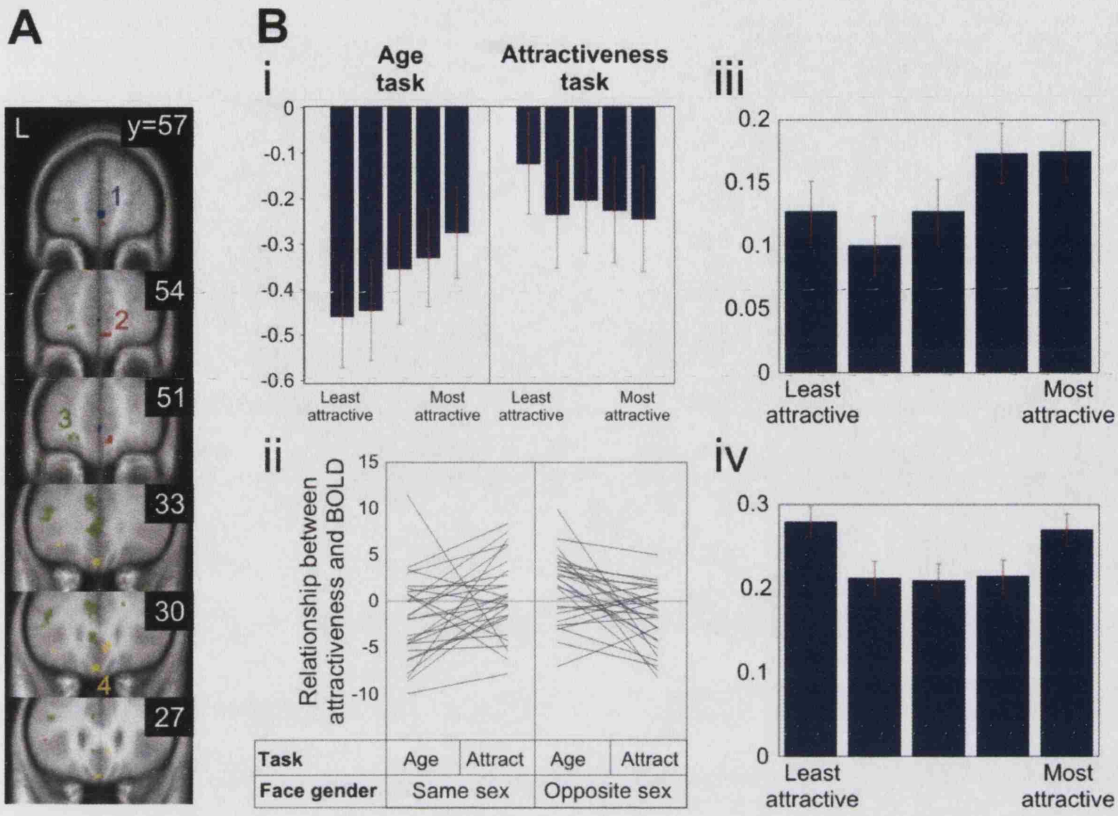

Figure 7.6: *OMPFC shows segregated responses to attractiveness*

a. Four different SPMs overlaid on mean structural image. Blue shows interaction between task and attractiveness. Red shows interaction of attractiveness, task, face gender and subject gender. Green shows main effect of attractiveness (attractive > unattractive) and yellow shows regions with significant positive second order effects of attractiveness. All thresholded at $p<0.001$ uncorrected. Peaks discussed in text and illustrated further in part (b) are labelled 1-4.

b. Illustration of effects shown in A:

   i. Inferior medial prefrontal cortex shows an interaction between task and attractiveness, with a linear positive increase in activation with increasing attractiveness under the age task but a decreasing response with increasing attractiveness under the attractiveness judgement task. Data derived as in Figure 7.3b.

   ii. Crossover interaction in inferior medial PFC between attractiveness, task, subject gender and face gender. More positive parameter estimates for the linear relationship between attractiveness and activation are obtained during the age task for opposite sex faces. Grey lines are individual subjects, blue is group mean.

   iii. A portion of medial OFC shows a main effect of attractiveness with greater responses to attractive than unattractive faces.

   iv. Posterior medial OFC shows significant non-linearity in response to attractiveness with extremes of attractiveness generating greater activation than faces of medium attractiveness.

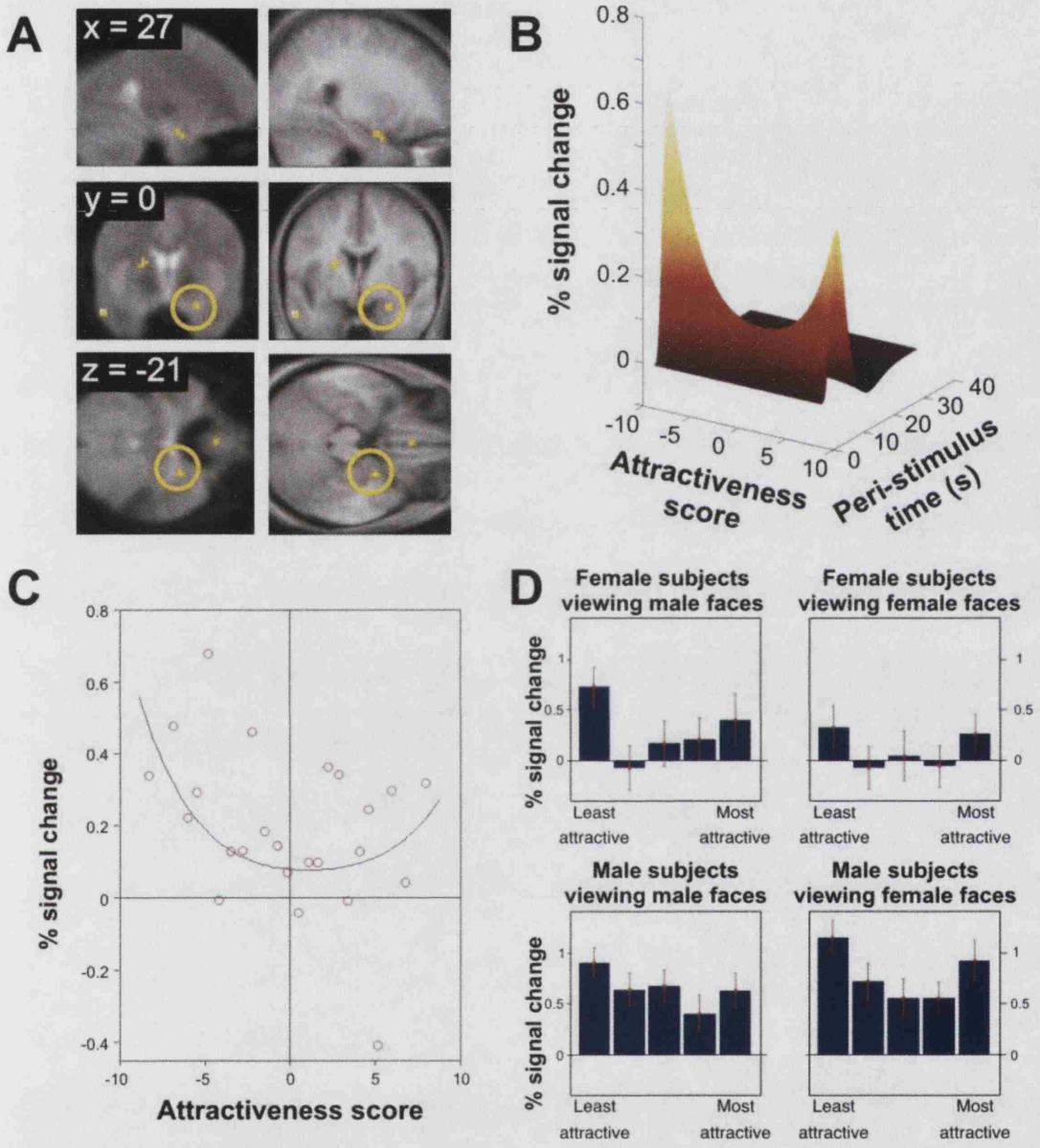

Figure 7.7: *Amygdala shows nonlinear responses across attractiveness space*

(Legend on following page)

Figure 7.7: *Amygdala shows nonlinear responses across attractiveness space*

a. Sagittal (top), coronal (middle) and axial (bottom) sections through group average scans with overlaid SPM of quadratic relationship between attractiveness score and BOLD activity. The statistical image is overlaid on the mean normalised EPI (left) and mean structural image (right) of the subject cohort. Activation is seen in anterior lateral right amygdala (x,y,z=27,0,-21; Z = 3.59; $p<0.05$ small volume corrected). SPM thresholded at $p<0.001$ uncorrected.

b. Mean fitted data from peak voxel. The plot was constructed using the mean parameter estimates of height and four parametric expansions describing the relationship between attractiveness score and BOLD and convolving the resulting average fit with SPM's canonical haemodynamic response function (HRF).

c. Mean responses in right amygdala voxels significant at $p<0.001$ uncorrected. The fitted data is derived as in B, the scatter (circles) shows the response profile from an unfitted model in which all 72 faces were modeled as individual event types. The resulting parameter estimates were arranged by mean attractiveness score, averaged in threes, and are plotted against attractiveness score.

d. Responses in right amygdala were non-linear, independent of subject and face gender. Female subjects (top row) show a non-linear relationship between activation and attractiveness when looking at male (left) or female (right) faces, as do male subjects (bottom row). Data from conjunction peak at x,y,z=27,0,-21 (Z=3.91, $p<0.05$ SVC), derived as in Figure 7.3b.

faces (Figure 7.7d), though this effect did not reach statistical significance. For the 16 subjects for whom arousal (pupillometric) data were available, inclusion of this data into the statistical model as a covariate of no interest made no discernable difference to the results in the amygdala (x,y,z = 27,0-24; Z = 3.64; $p<0.05$ SVC).

Discussion

In this study, I explored brain responses to faces which varied in attractiveness and additionally I examined task effects on those responses. The inclusion of a task that directly asked subjects to rate attractiveness along with a task that was unrelated (age judgement) allowed me to investigate the degree to which activations were task-dependent or -independent, along with the direct comparison of the profile of response in areas activated by the two tasks. By using a set of faces that varied on attractiveness, both linear and non-linear relationships between attractiveness and brain response were characterised.

The two primary questions addressed in this study (effects of task in reward paradigms and of non-linear stimulus-response profiles) have not been widely studied in humans. One early PET study has investigated the effect of judging attractiveness versus judging facial emotion, finding increased rCBF in left medial frontal cortex and left fronto-temporal junction when subjects made attractiveness judgements (Nakamura et al., 1998). A limitation of this study was a failure to optimise the range of attractiveness of the stimuli or to separate attractive and unattractive stimuli into separate blocks to look for an interaction between attractiveness and task. Studies comparing direct pleasantness judgements to other sorts of judgements in the olfactory domain have

similarly failed to test for interactions between reward value and judgement type (Zatorre et al., 2000; Royet et al., 2001).

A number of regions were more activated when judging attractiveness compared to age, including bilateral STS, bilateral posterior OFC extending into insula and medial prefrontal cortex. The activation of STS is of interest in that it has repeatedly been implicated in studies involving social evaluation. For example this region has been shown to respond to biological motion (Bonda et al., 1996; Puce et al., 1998; Puce and Perrett, 2003), face perception (Haxby et al., 2000), judgements of facial emotion (Narumoto et al., 2001; see also Chapter 4) and trustworthiness (Chapter 6), and theory of mind or attribution of intentions (Brunet et al., 2000; Castelli et al., 2000). The proposed role of some sectors of STS as an "intentionality detector" (Frith and Frith, 1999) ties these disparate functions together. I would suggest that judging attractiveness engenders such a process in that an integral component of assessing the attractiveness of a conspecific involves an invocation of another's intentions toward oneself.

In addition to STS, another region engaged by the attractiveness judgement task and previously implicated in social cognition was medial prefrontal cortex. Notably, medial prefrontal cortex has also been implicated in a recent study comparing evaluative judgements (e.g. I like George W. Bush: yes/no) to memory retrieval (Zysset et al., 2002) and a previous study in which subjects made attractiveness judgements (Nakamura et al., 1998). Moreover, a large body of literature implicates this region in theory-of-mind judgements and "self-referential" processing (e.g. Goel et al., 1995; Gallagher et al., 2000, 2002; Gusnard et al., 2001; Johnson et al., 2002; Kelley et al.,

2002). The current findings are consistent with an interpretation that judging attractiveness invokes a network that mediates self-referential processing in line with the idea that making an evaluative judgement implicitly invokes a representation of self.

Left amygdala also showed modulation by task with greater activity when subjects judged attractiveness than age. In other experiments in this thesis using similar manipulations, I have been unsuccessful in finding task differences in amygdala (Chapter 4 and 6) consistent with evidence that responses in this region are relatively automatic and independent of spatial attention (Vuilleumier et al., 2001; Bentley et al., 2003) or awareness (Morris et al., 1998b; Whalen et al., 1998; Williams et al., 2004). Other authors have found increases (Gorno-Tempini et al., 2001; Gur et al., 2002) or decreases (Critchley et al., 2000a; Hariri et al., 2000; Ochsner et al., 2002) in amygdala activity when an emotionally-directed task is compared to a control task. It is presently unclear what differences between paradigms and tasks might explain these discrepant results and this will be discussed at greater length in the general discussion.

The single region activated in the age judgement task (posterior cingulate/retrosplenial cortex) is a region consistently recruited in memory retrieval (Maguire, 2001a, b) including retrieval of memory for faces (Shah et al., 2001). This suggests that subjects used memory-dependent strategies for the age task, perhaps comparing the stimuli to familiar faces or by recalling their response on a previous encounter with the identical face (stimuli were repeated within each task).

Responses associated with attractiveness

Previous studies have suggested that areas of the brain that respond to other types of reward, such as monetary reward or pleasant chemosensory stimuli also respond more when observing attractive, relative to unattractive, faces (Aharon et al., 2001; O'Doherty et al., 2003b). These regions included medial orbitofrontal cortex and the nucleus accumbens in ventral striatum. The latter effect observed by Aharon et al (2001) in the context of a block-design study was not replicated by O'Doherty et al (2003b), who used a random stimulus order. O'Doherty et al (2003b) suggested that this discrepancy might be explained by a confound of reward anticipation in the block design. Despite the large number of subjects in both O'Doherty et al (2003b) and the current study, activation in ventral striatum in response to attractive faces was again not seen.

A surprising finding was that some activations previously defined in response to attractive faces in medial prefrontal cortex and medial OFC (O'Doherty et al., 2003b) were modulated by task. In the current study, greater responses were seen to attractive faces in medial PFC and to attractive opposite sex faces in medial OFC primarily in the age task. This was counter to the expectation that an explicit attractiveness judgement task might enhance activations in regions responsive to attractiveness. One possible explanation for this, given the large body of evidence highlighting these regions as responsive to reward, is that the unattractive faces become relatively rewarding in the attractiveness judgement task, perhaps because they are easier to classify. Importantly, other sectors of medial OFC showed task-independent linear relationships to attractiveness (Figure 7.6), so this explanation seems unlikely to be sufficient. Seemingly, some sectors of OFC show strongly stimulus-associated reward responses

whereas the response profiles of others are modulated by task context. In the context of instrumental conditioning, a dissociation has been observed between OFC regions responsive to reward processing and those associated with altering behavioural responses based upon experimental contingencies (O'Doherty et al., 2003a). Task-dependence in response to facial trustworthiness has also been observed in OFC in the study reported in Chapter 6 of this thesis.

A finding from a previous study (O'Doherty et al., 2003b) of greater activation to unattractive relative to attractive faces in lateral OFC and prefrontal cortex was not replicated in the current experiment. Such null results are hard to interpret. However, one possible explanation is a potential task difficulty confound in O'Doherty et al. (2003b) where the task used was gender judgement. This may have been more difficult to perform on the unattractive faces given that one determinant of high facial attractiveness is sexually dimorphic facial features (Perrett et al., 1998).

As in a previous study (O'Doherty et al., 2003b), relatively few regions showed effects dependent upon subject gender. However, one region which did was anterior cingulate cortex. Although this region showed a significant main effect of attractiveness (averaged across both subject genders) this effect was driven entirely by the male subjects as evidenced by a significant attractiveness-by-subject gender interaction (Figure 7.3). Interestingly, psychophysiological evidence consistent with such an effect was found in the pupillometry data available from a subset of subjects. This data showed a similar attractiveness-by-subject gender interaction with effects of attractiveness on pupil dilation expressed only in male subjects. The role of anterior cingulate cortex in generating and monitoring internal autonomic states is well-

characterised (Critchley et al., 2000c, 2003; Critchley, 2004; Teves et al., 2004), and these new data suggest that there are between-gender differences in arousal generated to attractive faces.

Non-linear response profiles

As predicted, amygdala showed a significant non-linear response profile, with greatest responses to attractive and unattractive faces, relative to those of medium attractiveness. These data extend our understanding of the role of the amygdala in social and emotional perception. Although earlier results have suggested that amygdala is not simply specialised to detect negatively-valenced stimuli or a specific type of emotion (e.g. Hamann et al., 1999, 2002; Garavan et al., 2001; Singer et al., 2004a, see also Chapter 4), to my knowledge there has been no previous description of a non-linear response profile. The demonstration of a non-linear response mode of this type is in keeping with the idea that the amygdala is tuned to the detection of events of emotional value, irrespective of valence, in the sensory environment (Dolan, 2002; Sander et al., 2003). Recently, it has been suggested that amygdala responses are driven by stimulus intensity (assumed to be a surrogate for "arousal") rather than valence (Anderson et al., 2003b; Small et al., 2003). Neither study supporting this conjecture included stimuli with valence close to zero, so it is hard to infer from these data that the amygdala is not representing aspects of valence. Moreover, in the current study, I am able to dissociate activity in the amygdala from at least one direct measure of arousal, pupil dilation. The inclusion of three aspects of pupillary responses as covariates of no interest in the statistical model in 16 subjects for whom these data were available had little effect on the statistical test for non-linearity, indicating that the quadratic relationship between

attractiveness rating and amygdala activation was not explained simply by greater arousal at extremes of attractiveness.

The finding of a portion of medial OFC with significantly non-linear responses to attractiveness, responding more to extremes than to faces of average attractiveness, is a strong demonstration of the utility of the approach adopted in the current study. Previous work (O'Doherty et al., 2003b) has shown a simple main effect of attractiveness in this region, and I note that in the current study, responses trended towards greater magnitude for attractive than unattractive faces. By using non-linear parametric expansions of the variable of interest in the current study, I am able to make stronger inferences about the underlying relationship than previous studies. Where linear associations are significant I can be more confident that they reflect linear relationships than in previous studies using parametric designs where higher order relationships were not modelled (e.g. O'Doherty et al., 2003b; see also Buchel et al., 1998a). In addition, this form of response to more and less positive stimuli relative to "average" is not unprecedented in this region: Elliott et al. (2003) demonstrated a similar second order relationship between monetary reward and BOLD activation in medial OFC. In that study, only five levels of reward were used, compared to 36 for both male and female faces in the current study, allowing a fuller characterisation here.

Summary

In this chapter, I have demonstrated that a number of brain regions are engaged when subjects judge facial attractiveness, and that these regions are largely dissociable from those that show relationships between the rating of attractiveness and BOLD activation.

With regard to the latter, a number, including reward-responsive regions in medial prefrontal and orbitofrontal cortex, showed modulation by the task that subjects were performing, though other areas did not show such task dependence. Along with the demonstration of non-linear response profiles in medial OFC this implies a complexity in the behaviour of these regions that has not previously been addressed by neurofunctional studies. Finally, the response profile of amygdala demonstrates a role for this region in encoding value (a non-linear function of valence) from stimuli in the environment.

Chapter 8: General discussion

The experiments described in this thesis comprise an exploration of one specific aspect of visual face perception, namely the processing of emotive qualities expressed in the human face. The experiments constitute a set of functional neuroanatomical studies of such processing and therefore this general discussion is structured around the observed components that are responsive during emotive face processing. An important theme that distinguishes between different components of this system reflects conditions under which individual components function: is their involvement automatic or better characterised by explicit or intentional processing? The discussion opens with a brief comment on what these different modes might represent. The individual components of the system for emotive face processing will then be discussed in turn, before some general critiques of the experimental approach are considered. Finally, I will return to a question that continues to be debated in the object processing literature, namely "are faces special?".

Modes of emotive face processing

Although generally designed to explore basic questions regarding the nature and location of coding for emotive aspects of the human face in the brain, a number of the experiments outlined address a subsidiary question, namely that of automaticity versus task-dependence in face processing. In the discussion of individual brain areas consistently activated in the studies within this thesis some tentative conclusions can be drawn about different modes under which these neuroanatomical components are engaged. The most important distinction is between those areas that activate

automatically in response to emotive stimulus characteristics and those that respond only under specific task requirements. A third category consists of those response profiles linked to stimulus characteristics, but modulated by task requirements. (For a perspective on the interpretation of different types of activation in the context of such mixed event-related and blocked designs as adopted here, see Donaldson, 2004).

Areas that show automatic responses to emotive characteristics (e.g. amygdala) are likely to be critical substrates in generating appropriate emotional responses to facial stimuli. By "appropriate", it is meant basic approach/withdraw or attend/ignore signals akin to emotions (preconscious dispositions or response patterns). It has been widely demonstrated that emotive aspects of faces are processed even when emotion is not a task-relevant variable – the critical evidence comes from brain imaging (e.g. Morris et al., 1996; Phillips et al., 1997; O'Doherty et al., 2003b), electrophysiological (e.g. Puce et al., 1999; Eimer and Holmes, 2002; Holmes et al., 2003) and behavioural investigations (e.g. Todorov and Uleman, 2002; Yamagishi et al., 2003). It is probable that areas that respond in such a task-independent fashion to emotive qualities in faces mediate some form of "online" monitoring of the (social) environment for threat or opportunity.

By contrast, regions showing task-dependent responses to emotive characteristics in faces, for example responding during a directed task in which an explicit judgement of emotional quality is required, are unlikely to play an integral role in implicit emotional responses to stimuli. Instead, I suggest that they are likely to reflect semantic-related processing relevant to the given task or, alternatively, comprise modules whose activity is related to task performance *per se*, perhaps in directing attention to task-relevant

aspects of the stimulus, or in generating a task-relevant output. Such a theoretical range of possibilities means that an activation may not have a unique interpretation. In such cases the most parsimonious explanations can be constrained, to some extent at least, by reference to the wider literature.

The final class of region, those showing stimulus-driven activations, but also exhibiting a degree of task-dependency (typically presenting as an interaction between task and stimulus characteristic) are open to a number of competing interpretations. Such a pattern (in the context of the task manipulations used in the current set of experiments) is unlikely to reflect a region responsible for automatic emotion processing in so far as the task manipulations adopted are likely insufficient to radically change early emotional processing of a stimulus[2]. Instead, I suggest that such a profile of interaction is likely to reflect a more complex role in emotive face processing and that allows a number of different interpretations, depending upon the nature of the interaction. Plausible interpretations are offered in each case.

Components of a system for emotive face processing

The basic components of the neural system for face processing have been extensively outlined (Haxby et al., 2000, 2002; Joseph, 2001; Grill-Spector, 2003; Palmeri and Gauthier, 2004). The two key components in basic visual processing of faces appear to be a region of posterior visual cortex in lateral occipital lobe and a region of the

[2] There is suggestion that complex cognitive strategies such as "reappraisal" or strong manipulations of spatial attention might be able to manipulate stimulus-driven emotional responses (Ochsner et al., 2002; Pessoa et al., 2002), but this seems unlikely to be the case for simple manipulations such as judging gender rather than emotion as adopted here.

fusiform gyrus in more medial occipito-temporal cortex. It is hypothesised that activity in the posterior region (sometimes known as "occipital face area" or OFA, but referred to here as face-responsive occipital region [FROR]) is evoked by structural encoding of faces or basic face/non-face discrimination. Fusiform face area (FFA) appears to be involved in higher order face processing, perhaps relating to components of identity processing (see also Chapter 3). However, studies of emotive qualities of faces such as those described in this thesis also demonstrate activation of fusiform cortex. This region is therefore included in the general discussion of areas subserving emotion processing. In contradistinction, FROR is rarely activated by emotive face processing, and is not discussed further.

The results of the studies reported in Chapters 3-7 are summarised in Figure 8.1. This shows peak voxels of activation, colour-coded into three categories according to the nature of the contrast that highlighted that voxel. There is a considerable degree of consistency in terms of activations within amygdala and fusiform cortex being almost exclusively found by effects relating to the basic emotional properties of the stimulus. STS shows two different forms of response profile: to the basic emotional properties of the stimulus or engaged by emotionally-directed tasks. Also apparent is the complex nature of responses within orbital and medial prefrontal cortex (OMPFC). This region is engaged sometimes by basic emotional properties, sometimes by emotionally-directed tasks and sometimes in an interaction between task and emotional quality. Each region is now discussed in turn.

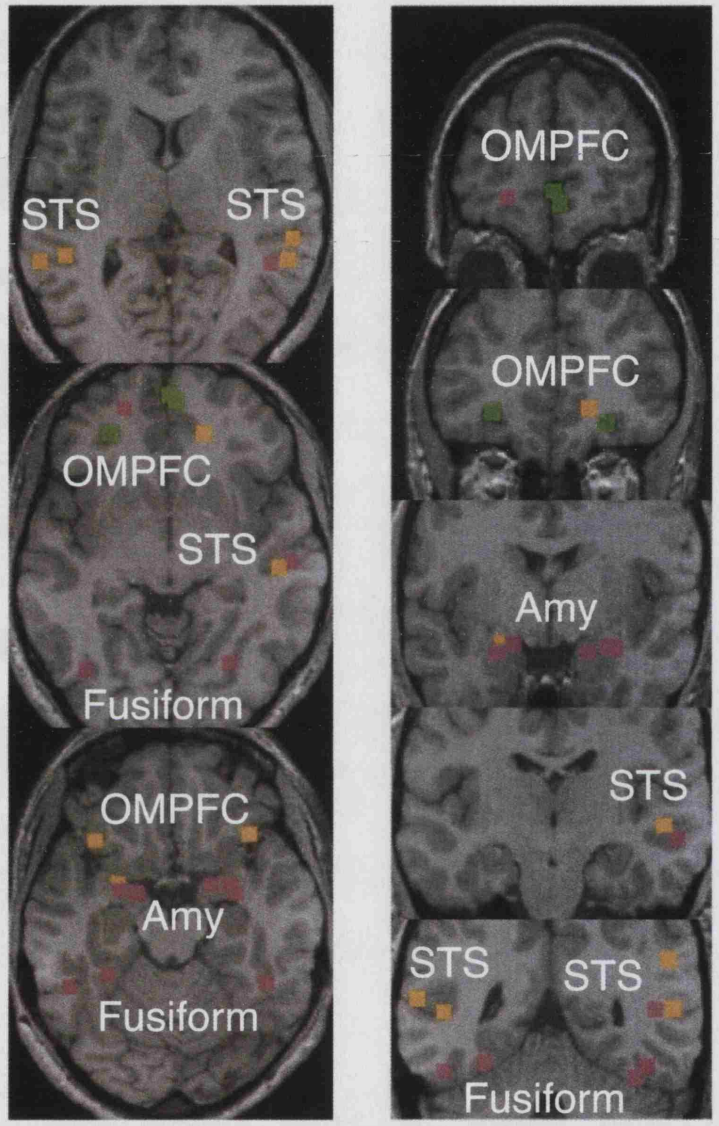

Figure 8.1: *Summary of results reported in this thesis*

This figure shows activations from Chapters 3-7 on a single subject's T1-weighted structural scan. Activations within amygdala ("Amy"), fusiform cortex, superior temporal sulcus (STS) or orbital and medial prefrontal cortex (OMPFC) are depicted as 7x7x7mm cubes centred at the peak coordinates and are colour-coded according the contrast that demonstrated activation. *Magenta* represents responses to basic emotional properties in the stimulus (facial emotion [Chapters 3-5], trustworthiness [Chapter 6], first or second order relationships with attractiveness [Chapter 7]). *Yellow* shows effects dependent upon task (greater responses during the emotion-directed task). *Green* areas are those that showed effects depending upon an interaction between the emotional variable and task, and are concentrated in OMPFC.

Fusiform cortex

Although fusiform cortex shows responses during basic face processing (Haxby et al., 1991, 1994; Sergent et al., 1992; Allison et al., 1994; Puce et al., 1995, 1996; Kanwisher et al., 1997), it has frequently been shown to manifest differential activity to emotional compared to neutral faces (Breiter et al., 1996; Dolan et al., 1996; Morris et al., 1998a; Vuilleumier et al., 2001, 2003a; Pessoa et al., 2002; Surguladze et al., 2003). The experiments reported in Chapters 4, 5 and 6 of this thesis replicate these general effects: in Chapter 4, high intensity expressions of four different basic emotions yielded greater activation in fusiform cortex than low intensity expressions. In Chapter 5, I showed that fear in low spatial frequency components of faces is sufficient to cause fusiform activation, and that this activation was independent of visual components of the stimulus used in making a gender discrimination. In Chapter 6, fusiform cortex showed responses linearly related to the perceived trustworthiness of a face, again an example of emotive qualities of the stimulus influencing responses in this region.

Data indicating fusiform responses to qualities other than identity in faces are problematic for theories that posit specialisation of fusiform for recognition of identity from faces (Haxby et al., 2000). However, a number of resolutions are possible:

1. Fusiform cortex may represent other aspects of faces alongside identity. Such a concept sits well with theories that FFA is not specialised for face processing per se, but for subordinate level categorisation of objects for which the viewer is expert (Gauthier et al., 1999, 2000a; Tarr and Gauthier, 2000). Thus FFA might respond to emotionality in faces as it subcategorises the emotional information in addition to performing its "normal" computations about the face.

2. Fusiform cortex might show enhanced activity to faces that display emotion as the presence of concomitant facial emotion increases the computational demands necessary to infer identity or subordinate level categorisation. Such an explanation fits well with perspectives from behavioural data showing lack of independence between identity and expression judgements (Schweinberger and Soukup, 1998; Schweinberger et al., 1999; de Gelder et al., 2003; Ganel and Goshen-Gottstein, 2004; see also the behavioural data presented in Chapter 3).
3. Enhanced fusiform activation to emotional faces might not represent increased intrinsic processing (or a bottom-up effect) in fusiform cortex but a modulatory effect, as a consequence of emotion processing, from other brain regions on earlier sensory processing of the face. This idea is supported by direct evidence from studies of functional connectivity between amygdala and fusiform or early extrastriate areas showing enhanced connectivity between the two under conditions in which the posterior regions respond to emotional faces (Morris et al., 1998a; Pessoa et al., 2002). It should be noted that the studies performed thus far use correlative techniques for assessment of connectivity (Friston et al., 1997), and cannot ascertain the direction of causation. A variant of this explanation is the idea that enhanced fusiform responses to emotive stimuli reflects enhanced attention to these stimuli.

Of the reasons for enhanced fusiform activation outlined above, the data in this thesis are most consistent with the last. In Chapter 3, no effect of repetition of emotional expression in face pairs was found in fusiform, a result that contrasts with the response profile in STS. If fusiform cortex coded specifically for emotional expressions (explanation 1 above), fMRI-adaptation to emotional expression should have been seen

in fusiform. In fact, fMRI-adaptation was only observed in fusiform to repetitions of identity, consistent with idea that fusiform cortex stores representations of facial identity. It has to be acknowledged that the inference that this is supportive evidence for feedback explanations of enhanced fusiform activity is based partly on a null result, and one should be cautious about such interpretations. However, more direct evidence in support of this hypothesis comes from a finding (described in Chapter 6) that fusiform regions show responses linearly related to the degree of perceived trustworthiness. Since the stimuli did not display overt emotion and there was inter-subject variability in rating trustworthiness, it seems more plausible that enhanced fusiform responses to untrustworthy faces were a result of feedback from anterior regions such as amygdala, that extract some valuation in relation to the stimuli, than the alternative explanations above.

A strong piece of evidence from data in the literature for top-down rather than stimulus-driven modulation of fusiform comes from studies in which the value of faces was learned rather than being an intrinsic quality. Conditioning studies show enhanced fusiform responses to fear-conditioned faces (Morris et al., 2001a; Armony and Dolan, 2002; Pizzagalli et al., 2003) and a recent study in which a set of faces was associated with different moral behaviours also demonstrated increased fusiform activity to emotive faces (Singer et al., 2004a). These studies counter-balanced the use of specific faces across subjects making a compelling case that it is the learnt emotional value of the stimuli that is reflected in fusiform responses (rather than intrinsic stimulus-driven responses). This concept is more compatible with the third possibility listed above than any of the alternatives. Finally, a recent study (Vuilleumier et al., 2004) in which temporal lobe epilepsy patients with and without amygdala involvement underwent

fMRI whilst viewing emotional and neutral faces provides strongest evidence to date of a feedback mode of fusiform engagement. Patients with sclerosis involving amygdala showed less modulation of fusiform response by emotion than patients without amygdala involvement or healthy controls. This was not due to generally abnormal fusiform responses, as the normal enhancement seen when spatial attention is directed to faces was still present in the patients with amygdala damage.

The conclusion I reach is that enhanced fusiform activity to emotive stimuli results from feedback from anterior regions encoding the emotional value of the stimulus. Consistent with this idea, a recent intracranial recording study found later emotional modulation of ERP responses in fusiform regions than amygdala, suggestive of a temporal hierarchy in response profiles (Krolak-Salmon et al., 2004). Similarly, responses associated with decoding of emotional information are reported as occurring relatively late in inferotemporal cortex neurons (Sugase et al., 1999). A modulatory effect must of course be mediated by anatomy and it has been shown that there are reciprocal connections between amygdala and inferotemporal regions (a putative homologue of human FFA) in macaque monkeys (Amaral and Price, 1984; Amaral et al., 1992).

Amygdala

Although it has been known for some time that the amygdala is important in social and emotional processing (Weiskrantz, 1956), a more precise characterisation of its functions has recently emerged. Humans with bilateral amygdala damage show impairments in recognition of emotion from faces, often limited to fear and/or one other

negative emotion (Fine and Blair, 2000). The amygdala is not exclusively associated with fear processing: a range of experiments in non-human primates demonstrate roles for the amygdala in positive affective processing, both of non-face (e.g. Rolls, 1999; Ono and Nishijo, 2000) and face stimuli (Nakamura et al., 1992). However, the relative selectivity of recognition deficits to fearful faces and the indication from PET data that human amygdala responds increasingly as faces move away from happy through neutral to fearful expressions (Morris et al., 1996) has led to an assumption that the human amygdala is specialised for fear perception. Evidence from the experiments described in this thesis, and from other results in the literature, offer a broader perspective.

In Chapter 6, I presented evidence that the amygdala response to male faces varied as a function of the perceived trustworthiness of the face, with greatest responses to the least trustworthy individuals. This indicates a broader role for the amygdala in face perception than fear recognition and is consistent with data from patients with bilateral amygdala damage (Adolphs et al., 1998). Minimally, this result indicates that the amygdala monitors a broader category of social threat than expressions of fear. Moreover, recent results indicate that faces representing morally good or bad behaviours activate amygdala to a greater extent than neutral faces (Singer et al., 2004a). This extends the result from Chapter 6 and will be discussed in more detail below.

In Chapter 4, I presented evidence that the amygdala responds to a variety of facial expressions, relative to neutral faces, again consistent with a broad role for the amygdala in emotive face perception. Finally, in a finding predicted from the results reported in other chapters, data presented in Chapter 7 indicated that the right amygdala

showed a quadratic response profile to faces varying on attractiveness, with greatest responses to attractive and unattractive faces relative to those of medium attractiveness. This latter finding presents a broader characterisation on what aspects of faces the amygdala encodes.

As mentioned in the general introduction, there are a number of theories that purport to explain the functions of the human amygdala in a unitary manner. The consistency of these data in this thesis with respect to each of these will now be discussed. One model, already mentioned, is that the human amygdala is a fear module. The data presented in this thesis are incompatible with such an explanation. Although Chapter 6 demonstrated enhanced amygdala responses to socially threatening faces, Chapters 4 and 7 both include data that are inconsistent with the concept of the amygdala as a fear-specific unit. Chapter 4 showed amygdala responses to a range of emotions (including fear, but additionally happiness, disgust and sadness) when the expression was of high intensity. Chapter 7 showed amygdala responses to faces that were attractive or unattractive, again inconsistent with the idea of an amygdala specialisation for fear, or even for negatively valenced stimuli.

An alternative to the fear specialisation hypothesis is the proposal that some sectors of the amygdala act as an ambiguity detector, responding when more information is needed to interpret sensory input from the environment, and commanding attentional resources to decrease this ambiguity (Whalen, 1998; Davis and Whalen, 2001). This hypothesis receives some support from the experiments described in this thesis, but in my view it is not the most parsimonious explanation for the data. For example, it could be the case that observing attractive and unattractive individuals is associated with

ambiguity and a need for more information, but it seems more probable that such responses are associated with decoding the relevance or value of such faces (see below).

Most recently it has been proposed that the amygdala is a relevance detector (Sander et al., 2003). This hypothesis suggests that the amygdala is not solely activated by positive or negative stimuli but by those features in the environment which are most relevant to adaptive behaviour. The data from Chapters 4 and 7 are consistent with this hypothesis: facial expressions of emotion are relevant features for behaviour regardless of which specific emotion is exhibited (Chapter 4). Similarly, both attractive and unattractive faces (Chapter 7) are worthy of attention (albeit for different reasons from one another). The result from Chapter 6 that the amygdala responds more to untrustworthy than trustworthy faces is less obviously compatible with this proposal – it seems surely the case that trustworthy individuals are "relevant". One possibility is that this result reflects a default mode of activity with respect to unfamiliar faces. Where faces are familiarised by means of an interactive game, the amygdala shows responses to both trustworthy and untrustworthy individuals (Singer et al., 2004a).

fMRI does not have sufficient resolution to assign responses to specific subnuclei of the amygdala. Thus, it should be borne in mind that the responses to distinct facial emotions or to extremes of attractiveness may represent the activity of distinct subsystems within the amygdala. fMRI-adaptation based designs (such as that in Chapter 3) may be able to characterise the underlying code (see Naccache and Dehaene (2001) for an example outside the realm of emotion). The problem with such an experiment is that some hypotheses predict a null result, and might therefore be untestable using classical inference. Similar to the limits on spatial resolution that

preclude delineation of subnuclear responses, fMRI is limited in the temporal domain in that it detects a haemodynamic correlate of aggregate neural activity over a period of time (Logothetis et al., 2001). Thus, differences between categories that fMRI cannot detect may exist at the temporal level (e.g. early responses to some emotions and later to others).

One interesting issue addressed in a number of the experiments in this thesis is the response of the amygdala to task manipulations. In experiments in Chapters 4, 6 and 7, where this was explicitly tested, responses were found independent of task. However, in the experiment described in Chapter 7 a greater response was seen in left amygdala during attractiveness judgements compared to age judgements. As stated in Chapter 7, it is not immediately apparent what factors might be responsible for the discrepancies between studies that find increases in amygdala activity with emotionally directed tasks (e.g. Gorno-Tempini et al., 2001; Gur et al., 2002), decreases in amygdala activity (e.g. Critchley et al., 2000a; Hariri et al., 2000; Ochsner et al., 2002) and those that find no difference (e.g. Chapters 4 and 6). The general consensus from the experiments in this thesis, where I would argue that the experimental designs were better controlled than for many of the above studies, was an absence of task effects in the amygdala. Such a result is compatible with theories that emphasise the role of the amygdala in *automatic* evaluation of stimuli in the environment (Dolan and Vuilleumier, 2003).

Superior temporal sulcus

The discovery of face cells in monkey STS (Perrett et al., 1982) was the first indication that this area of cortex might include specialisations for social processing. Subsequent

single unit recording studies and human brain imaging experiments have revealed a wide range of functions that engage STS. There is some suggestion that the face cells in STS are specialised for processing different features from inferotemporal regions, for example facial expression (Hasselmo et al., 1989), or eye-gaze (Perrett et al., 1985) or head direction (Perrett et al., 1991; Eifuku et al., 2004) rather than identity in IT cortex (Hasselmo et al., 1989; Eifuku et al., 2004). Similar dissociations are evident in data from studies in humans (e.g. Haxby et al., 2000; Hoffman and Haxby, 2000; Narumoto et al., 2001). In Chapter 3 of this thesis, evidence was presented that a portion of STS showed fMRI adaptation to repetitions of facial expressions whereas fusiform showed adaptation to repetitions of identity (across different stimuli). This is consistent with the dissociation described above. However, a finessing of this position is warranted in that a more posterior sector of STS showed adaptation to repeated identity (along with weaker adaptation to repeated expression). This is consistent with the recent suggestion that identity-processing cells should be found amongst regions specialised for other aspects of face processing, but not necessarily the converse (Tiberghien et al., 2003).

More generally, the STS showed activations to high intensity relative to low intensity expressions of emotion (Chapter 4), in making judgements about emotion rather than gender (Chapter 4) and social judgements about trustworthiness or attractiveness compared to judgements about age (Chapters 6 and 7 respectively). In addition, the STS shows responsiveness to static faces (Puce et al., 1996; Kanwisher et al., 1997), biological motion (Puce and Perrett, 2003), and during tasks that require mentalising (Frith and Frith, 1999). This assortment of responses suggests that the function of this region extends beyond registering biological motion or implied motion (Allison et al., 2000). Instead, as repeatedly suggested in this thesis, these activations are consistent

with the proposed role of STS in "intention detection" (Frith and Frith, 1999). The response profile of this region differs from those in amygdala in that (with the exception of facial expressions) responses dependent upon the nature of the stimulus were not seen. Instead, enhanced activation was observed when participants made explicit judgements about social characteristics of stimuli rather than less socially-directed judgements. This suggests that this area exhibits less automaticity than amygdala, being activated only under specific task conditions. The exception of facial expressions (where an additive effect of task and stimulus was found) is probably a result of specific representations of facial emotion in this area, as shown in single unit recordings in monkeys (Hasselmo et al., 1989) and Chapter 3.

Orbital and medial prefrontal cortex

Although Phineas Gage suffered the accident that resulted in damage to OMPFC, and its attendant personality change almost 150 years ago the specific role for this region in social and emotional cognition have only recently been delineated. These were reviewed briefly in the Introduction to this thesis, and have been reviewed extensively in recent literature (Brothers, 1990; Rolls, 1996, 2000; Bechara et al., 2000; Schultz et al., 2000b). The results described in this thesis are consistent with the early proposal of Brothers (1990) that the OMPFC (nowadays taken to include ventral and medial sectors of prefrontal cortex) subserves a specific function in processing social stimuli.

As befits a region of prefrontal cortex, the particular roles of OMPFC in social processing appear more complex than those of posterior regions of the brain. For example, in the experiment described in Chapter 5, lateral OFC was shown to respond

to fear in low spatial frequency (LSF) components of faces, but only when the gender of the face was reported from those components (during overt recognition). This can be contrasted with fusiform cortex which showed automatic responses to the LSF fear components independent of subjective report. Strikingly, the coordinates of the activated peak were nearly identical to that reported by Vuilleumier et al (2002) where a response was seen only when a patient with hemi-spatial neglect reported seeing (rather than neglecting) a fearful face in a visual hemifield. These results indicate that some areas within OMPFC represent an interaction between emotional stimulation and conscious perception.

Similar to STS, sectors of OMPFC showed main effects of task in the experiments described in Chapters 4 and 7. In Chapter 4 a large area of ventromedial prefrontal cortex showed enhanced responses when subjects made overt judgements of emotion rather than gender, and in Chapter 7, posterior OFC extending into insula showed activity when subjects judged attractiveness rather than age. However, other areas of OFC showed stimulus-dependent responses (e.g. medial OFC in Chapter 7, whereby different regions showed linear and nonlinear responses to attractiveness) or responses dependent upon an interaction between stimulus and task (e.g. lateral OFC in Chapter 6 or medial regions in Chapter 7). In other words, in these latter experiments regions with OFC showed responses of magnitude that depended upon the stimulus and upon the task required of the subject. In this regard, OMPFC showed responses a degree more complex than other regions repeatedly activated in the experiments reported in this thesis. Figure 8.1 demonstrates clearly that responses within OMPFC were of a varied nature. Note in particular, that it was the only region that showed such interactions between stimulus and task.

General theories of prefrontal function emphasise the flexible nature of output responses mediated by this region (Duncan, 2001; Funahashi, 2001), with perhaps regional specialisation for preferred types of information. There is a consensus that inferior and medial regions of prefrontal cortex are specialised for emotional or motivational functions whereas dorsal and lateral regions are implicated in processing "cold" cognition. Within these divisions, it is currently unclear how consistently sub-specialisations are found. One hypothesised sub-specialisation is that of a dissociation between representation of reward and punishment in medial and lateral OFC respectively (O'Doherty et al., 2001a). Within the experiments described in this thesis there is relatively little spatial agreement in response profiles of a similar nature. For example, a task by face-type interaction in experiments described in Chapters 6 and 7 occur in different regions. Similarly, the areas expressing main effects of task in Chapters 4 and 7 were spatially discontiguous. This lack of segregation of apparently similar functions is consistent with a suggestion that OFC is not strongly spatially segregated but consists of flexible neuronal populations whose behaviour adapts contextually. The lack of evidence for strong segregation notwithstanding, it is important to note that cytoarchitectonic and connectivity differences do exist within ventral prefrontal cortex (Barbas and Pandya, 1989; Morecraft et al., 1992; Hof et al., 1995; Price et al., 1996), which conceivably might reflect underlying functional differences.

Limitations of the experimental approach

Some general limitations of the experiments reported in this thesis need to be acknowledged. Firstly, there are a number of general limitations with BOLD fMRI.

These include a limited understanding of the relationship between neurophysiology and BOLD responses. This limits the interpretation of BOLD activation studies – for example, both inhibitory and excitatory processing can lead to net increases in BOLD signal since both increase net metabolic demands. This fact does not have the detrimental implications that some critics of functional neuroimaging assume. It is often assumed that recording from single neurons constitutes a gold standard for interpreting a neural code. However, this relies on the assumption that single units provide a fundamental level of description for encoding or representing cognition. Instead, it can reasonably be argued that complex cognition is an emergent property of large scale neural networks as opposed to the micoresolution of the single unit.

A second problem with BOLD fMRI is an inhomogeneity in brain coverage. Regions close to air or bone boundaries, for example in ventral aspects of the brain, suffer inhomogeneities in magnetic field which produce signal distortion and dropout. This reduces signal quality in affected regions including ventral temporal cortex, orbitofrontal cortex, brainstem and ventral midbrain. Steps to reduce this effect were adopted in some of the experiments described in this thesis: the use of relatively low field strengths and the use of tilted acquisition in the experiment described in Chapter 7 assisted reconstruction in ventral frontal regions (Deichmann et al., 2003). Notably, this limitation of fMRI does not affect inference when significant results are found, it merely increases the likelihood of false negatives. This comprises another reason that null results in fMRI are difficult to interpret.

A third potential problem with BOLD fMRI is the fact that statistical inference using informed basis sets, such as the haemodynamic response function (HRF) as adopted

throughout this thesis, may miss response profiles that differ from a canonical response profile. It has been demonstrated that different brain regions do show different HRF shapes (Miezin et al., 2000), and the use of a canonical HRF in the statistical models used for data analysis in this thesis may potentially limit sensitivity to effects in brain regions with haemodynamic response profiles different from the canonical HRF. It is possible that certain subcortical nuclei along with brain stem and mid-brain, that might have been expected to activate in studies concerning emotion (Panksepp, 2004), possess different HRF properties compared to other regions that did show activation.

Beyond potential problems with BOLD fMRI, there are limitations to the statistical approach adopted. Specifically, in a number of experiments in this thesis, null results are reported (e.g. no interaction between task and stimulus type in amygdala in Chapters 4, 6 and 7). It is of course impossible to accept the null hypothesis of no effect, and it should be remembered that when discussing such null results (e.g. the inference that the amygdala is relatively task-insensitive) that a more sensitive analysis or technique might have yielded a positive result. Wherever such results are reported in this thesis, it is in the context of a significant main effect and frequently reinforced by presentation of data from conjunction analysis across the simple effects indicating an effect in both of the simple effects.

A methodological issue facing the more complex parametric designs of Chapters 6 and 7 is the assumption that ratings of faces in the scanner were matched by the post-hoc parametric ratings acquired in debriefing. This limitation is enforced by the difficulty of getting fully parametric ratings inside the scanner within the timeframe required for efficient experimental design. A number of factors might effect a change in complex

social ratings such as trustworthiness and attractiveness between rating contexts. For example, familiarity with the stimuli is increased by the time of debriefing, and familiarity has been demonstrated to relate to perceived attractiveness (Peskin and Newell, 2004). In the experiment described in Chapter 7 the assumption that ratings inside the scanner were matched by those outside was empirically tested and found to be true. (A similar analysis was not possible for the data concerning trustworthiness as subjects rated only half of the stimuli inside the scanner by experimental design.) To a certain degree, the fact that significant results were found based upon these post-scanning ratings itself supports the contention that such ratings were adequate to assess responses to the stimuli.

Another possible criticism of some experiments in this thesis is the equivalence assumed between *automaticity* and task-independence for the limited number of tasks tested. The suggestion that amygdala responses to some emotive faces are automatic is however supported by a wealth of data from other experiments that utilise a variety of paradigms such as task manipulations, spatial attention, visual masking, and blindsight (e.g. Whalen et al., 1998; Morris et al., 1998b, 2001b; Vuilleumier et al., 2001; Anderson et al., 2003a; reviewed in Dolan and Vuilleumier, 2003). It should be remembered that the suggestion in this thesis that such automaticity applies to more complex judgements such as attractiveness and trustworthiness has thus far only been demonstrated for simple manipulations of task and not for the same range of paradigms as for emotional or fear-conditioned faces.

Are faces special?

An ongoing debate exists in the literature as to whether faces are privileged visual objects with a neural architecture devoted to their processing (Kanwisher, 2000; Tarr and Gauthier, 2000). The experiments described in this thesis explored a specific feature of face processing, namely facial aspects that are emotive. The debate as to whether faces are treated as special visual objects has primarily centred on one region within occipito-temporal cortex – the FFA. However, in the experiments in this thesis a number of regions beyond FFA were activated during face processing. These included STS, amygdala and OMPFC. One aspect of activity in these regions addressed in this thesis is the degree to which emotion-related activity depends upon emotion-directed processing. In the case of the amygdala, and arguably to some degree in fusiform cortex, STS, and OMPFC, processing has been shown to be relatively automatic, with responses to emotional qualities in faces being expressed even when the emotional variable is not task-relevant. This suggests an additional manner in which faces are special visual objects: they contain emotional content that is automatically processed by the brains of observers and can engender emotion (perhaps also automatically) in an observer.

Does this attribute of neural processing for faces make them special? The answer to this question depends upon one's understanding of the term "special". In one sense of the term, namely a requirement that those areas responsive to faces must respond exclusively to such stimuli, the answer is no. With the possible exception of fusiform cortex, all of the brain regions highlighted are activated in a range of affective processing in different sensory modalities. Posterior STS has been implicated in

psychological processes loosely connected by an underlying involvement in "intention detection" (Frith and Frith, 1999). The amygdala has been shown to be activated by rewarding or aversive stimuli in multiple modalities (Zald, 2003) as has OMPFC (Kringelbach and Rolls, 2004). Thus, activations during visual face processing are not special in the stringent sense of exclusive activation to a single class of stimulus. But the frequency of activation in these regions in studies in this thesis does indicate that these regions are interested in emotive aspects of faces, and that processing of such aspects is frequently automatic. This is consistent with faces being "special" in an alternative sense of the word.

What are the implications of faces being special visual objects by virtue of their engagement of emotional systems in the brain? At one level, this conclusion merely serves to remind us how important faces are to humans, and by extension to other mammalian species which possess cortical circuitry responsive to faces (e.g. non-human primates Perrett et al., 1982; and sheep Kendrick and Baldwin, 1987). However, there are two further implications of importance. The first is that the array of extra-visual responses emphasise the multi-faceted nature of facial stimuli: they are not like other visual stimuli where identification represents a singular goal of object processing. With faces, an assortment of post-identification[3] processes are important, or might be important in certain contexts: is this face showing emotion, is that emotion positive or negative? Where is this person looking? Is this individual to be trusted? The brain has

[3] "*Post*-identification" may not be an ideal term, as it suggests a temporal order or serial process. Instead it is likely that these emotive qualities are processed in parallel (Bruce and Young, 1986), and there is some evidence that facial emotion, for example, is processed more quickly than identity (compare Eimer, 2000; Schweinberger et al., 2002; Eimer et al., 2003)

evolved to assess all these attributes and many more *without a necessary dependence on conscious control.*

A second implication of the close association between emotional systems and faces might pertain to the developmental trajectory of face processing. There is evidence that newborn infants respond to faces extremely early in life (Goren et al., 1975; Johnson et al., 1991), despite poor visual acuity (Atkinson et al., 1979). It has been suggested that early vision and visual attention is largely supported by subcortical pathways via superior colliculus rather the cortical route (Johnson and Morton, 1991). Notably, a growing body of evidence implicates this same pathway in adult face processing, particularly of emotive faces (emotionally expressive or fear-conditioned, Morris et al., 1998b, 1999, 2001b; Vuilleumier et al., 2003a; Pasley et al., 2004). It seems plausible that these two aspects of face processing are two sides of the same coin: the infant's attention to faces is driven by a rudimentary internal representation of what is of importance while the adult brain adopts the same pathway to generate an early signal that attention should be paid to an emotive face stimulus.

One of the implications of an association between developmental aspects of face processing and rapid processing of emotional faces in adulthood is the meaning of face processing deficits and amygdala abnormalities reported in patients with developmental disorders including autism. Although both enlargements and reductions in size of the amygdala have been reported in structural studies of the autistic brain (Abell et al., 1999; Aylward et al., 1999; Howard et al., 2000; Pierce et al., 2001), a recent developmental study (Schumann et al., 2004) suggested that the best summary of the structural amygdala abnormalities in autism is that the developmental trajectory is

abnormal. Additional evidence for amygdala involvement in autism comes from behavioural (Adolphs et al., 2001) and functional imaging (Baron-Cohen et al., 1999) studies, as summarised in Baron-Cohen et al (2000). In particular, the processing of faces appears disrupted in autism, with autistic children showing less interest in faces (Pedersen et al., 1989) and abnormal responses to eye-gaze direction (Senju et al., 2003; Senju et al., 2004) in addition to the impairment in recognition of emotive features (Adolphs et al., 2001). A number of functional imaging studies have demonstrated abnormal neural responses to faces in autistics (e.g. Schultz et al., 2000a; Pierce et al., 2001), primarily in occipito-temporal regions including fusiform gyrus, but additionally in amygdala. One hypothesis might be that autistic children have a deficit in a subcortical pathway suggested to underpin the development of orienting to faces (Johnson and Morton, 1991) and adult responses to emotive faces under conditions of restricted conscious awareness (Morris et al., 1999; Vuilleumier et al., 2003a; Pasley et al., 2004; Williams et al., 2004; Chapter 5). (For a similar perspective on the development of face processing in autism, see Grelotti et al., 2002). Some recent evidence consistent with such a suggestion is the finding that autistic subjects show poorer performance on a range of face processing tasks when using low spatial frequency cues in faces than healthy controls (Deruelle et al., 2004).

Conclusion

The experiments in this thesis explored how the brain processes emotive qualities in human faces. I have established that these qualities are processed by brain regions partly dissociable from those involved in face detection and identity processing. My studies highlight the fact that a key structure in emotive face processing is the

amygdala, which shows task-independence in its function. I have shown that automaticity for emotive face processing is not limited to relatively simple characteristics, such as facial expressions of basic emotions, but extends to more complex psychological constructs, such as trustworthiness and attractiveness. A set of regions beyond the amygdala are invoked by explicit task requirements directing attention to emotive characteristics in the face: which specific regions appears to vary by the specific quality in question, but STS is frequently invoked by such explicit social judgements, supplementing its role in representation of specific facial emotions. Finally, I have demonstrated that ventral sectors of prefrontal cortex show processing of emotive qualities but with activity patterns modulated by task requirements, perhaps reflecting a role in linking social judgements to explicit behavioural patterns.

Publications arising from the work described in this thesis

Winston, J. S., and Dolan, R. J. (2004). Feeling States in Emotion: Functional Imaging Evidence. In Feelings and Emotions: The Amsterdam Symposium, A. S. R. Manstead, N. Fridja, and A. Fischer, eds. (Cambridge, Cambridge University Press), pp. 204-220.

Winston, J. S., Henson, R. N., Fine-Goulden, M. R., and Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. J Neurophysiol *92*, 1830-1839.

Winston, J. S., O'Doherty, J., and Dolan, R. J. (2003). Common and distinct neural responses during direct and incidental processing of multiple facial emotions. Neuroimage *20*, 84-97.

Winston, J. S., Vuilleumier, P., and Dolan, R. J. (2003). Effects of low-spatial frequency components of fearful faces on fusiform cortex activity. Curr Biol *13*, 1824-1829.

Winston, J. S., Strange, B. A., O'Doherty, J., and Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. Nat Neurosci *5*, 277-283.

Bibliography

Abell, F., Krams, M., Ashburner, J., Passingham, R., Friston, K., Frackowiak, R., Happe, F., Frith, C., and Frith, U. (1999). The neuroanatomy of autism: a voxel-based whole brain analysis of structural scans. Neuroreport *10*, 1647-1651.

Adams, R. B., Jr., Gordon, H. L., Baird, A. A., Ambady, N., and Kleck, R. E. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. Science *300*, 1536.

Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature *372*, 669-672.

Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. R. (1995). Fear and the human amygdala. J Neurosci *15*, 5879-5891.

Adolphs, R., Damasio, H., Tranel, D., and Damasio, A. R. (1996). Cortical systems for the recognition of emotion in facial expressions. J Neurosci *16*, 7678-7687.

Adolphs, R., Tranel, D., and Damasio, A. R. (1998). The human amygdala in social judgment. Nature *393*, 470-474.

Adolphs, R. (1999). Social cognition and the human brain. Trends Cogn Sci *3*, 469-479.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., Anderson, A., Lee, G. P., and Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. Neuropsychologia *37*, 1111-1117.

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., and Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. J Neurosci *20*, 2683-2690.

Adolphs, R. (2001). The neurobiology of social cognition. Curr Opin Neurobiol *11*, 231-239.

Adolphs, R., Sears, L., and Piven, J. (2001). Abnormal processing of social information from faces in autism. J Cogn Neurosci *13*, 232-240.

Adolphs, R. (2002a). Recognizing Emotion From Facial Expressions: Psychological and Neurological Mechanisms. Behav Cogn Neurosci Rev *1*, 21-61.

Adolphs, R. (2002b). Neural systems for recognizing emotion. Curr Opin Neurobiol *12*, 169-177.

Adolphs, R., Baron-Cohen, S., and Tranel, D. (2002). Impaired recognition of social emotions following amygdala damage. J Cogn Neurosci *14*, 1264-1274.

Adolphs, R. (2003a). Is the human amygdala specialized for processing social information? Ann N Y Acad Sci *985*, 326-340.

Adolphs, R. (2003b). Cognitive neuroscience of human social behaviour. Nat Rev Neurosci *4*, 165-178.

Aguirre, G. K., Zarahn, E., and D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. Neuroimage *8*, 360-369.

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., and Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. Neuron *32*, 537-551.

Allison, T., Ginter, H., McCarthy, G., Nobre, A. C., Puce, A., Luby, M., and Spencer, D. D. (1994). Face recognition in human extrastriate cortex. J Neurophysiol *71*, 821-825.

Allison, T., Puce, A., Spencer, D. D., and McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. Cereb Cortex *9*, 415-430.

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. Trends Cogn Sci *4*, 267-278.

Amaral, D. G., and Price, J. L. (1984). Amygdalo-cortical projections in the monkey (Macaca fascicularis). J Comp Neurol *230*, 465-496.

Amaral, D. G., Price, J. L., Pitkanen, A., and Carmichael, S. T. (1992). Anatomical Organization of the Primate Amygdaloid Complex. In The Amygdala: Neurobiological

Aspects of Emotion, Memory, and Mental Dysfunction, J. P. Aggleton, ed. (New York, Wiley-Liss), pp. 1-66.

Anderson, A. K., and Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. Nature *411*, 305-309.

Anderson, A. K., and Phelps, E. A. (2002). Is the human amygdala critical for the subjective experience of emotion? Evidence of intact dispositional affect in patients with amygdala lesions. J Cogn Neurosci *14*, 709-720.

Anderson, A. K., Christoff, K., Panitz, D., De Rosa, E., and Gabrieli, J. D. (2003a). Neural correlates of the automatic processing of threat facial signals. J Neurosci *23*, 5627-5633.

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., Gabrieli, J. D., and Sobel, N. (2003b). Dissociated neural representations of intensity and valence in human olfaction. Nat Neurosci *6*, 196-202.

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., and Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. Nat Neurosci *2*, 1032-1037.

Armony, J. L., and Dolan, R. J. (2002). Modulation of spatial attention by fear-conditioned stimuli: an event-related fMRI study. Neuropsychologia *40*, 817-826.

Atkinson, J., Braddick, O., and French, J. (1979). Contrast sensitivity of the human neonate measured by the visually evoked potential. Investigative Ophthalmology and Visual Science *18*, 210-213.

Aylward, E. H., Minshew, N. J., Goldstein, G., Honeycutt, N. A., Augustine, A. M., Yates, K. O., Barta, P. E., and Pearlson, G. D. (1999). MRI volumes of amygdala and hippocampus in non-mentally retarded autistic adolescents and adults. Neurology *53*, 2145-2150.

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., and Hyde, J. S. (1992). Time course EPI of human brain function during task activation. Magn Reson Med *25*, 390-397.

Barbas, H., and Pandya, D. N. (1989). Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. J Comp Neurol *286*, 353-375.

Bard, P. (1928). A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system. Am J Physiol *84*, 490-515.

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? Cognition *21*, 37-46.

Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., and Williams, S. C. (1999). Social intelligence in the normal and autistic brain: an fMRI study. Eur J Neurosci *11*, 1891-1898.

Baron-Cohen, S., Ring, H. A., Bullmore, E. T., Wheelwright, S., Ashwin, C., and Williams, S. C. (2000). The amygdala theory of autism. Neurosci Biobehav Rev *24*, 355-364.

Bartels, A., and Zeki, S. (2000). The neural basis of romantic love. Neuroreport *11*, 3829-3834.

Bartels, A., and Zeki, S. (2004). The neural correlates of maternal and romantic love. Neuroimage *21*, 1155-1166.

Baxter, M. G., and Murray, E. A. (2002). The amygdala and reward. Nat Rev Neurosci *3*, 563-573.

Baylis, G. C., Rolls, E. T., and Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. Brain Res *342*, 91-102.

Bechara, A., Damasio, H., and Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. Cereb Cortex *10*, 295-307.

Bentin, S., Allison, T., Puce, A., Perez, E., and McCarthy, G. (1996). Electrophysiological studies of face perception in humans. Journal of Cognitive Neuroscience *8*, 551-565.

Bentley, P., Vuilleumier, P., Thiel, C. M., Driver, J., and Dolan, R. J. (2003). Cholinergic enhancement modulates neural correlates of selective attention and emotional processing. Neuroimage *20*, 58-70.

Bird, C. M., Castelli, F., Malik, O., Frith, U., and Husain, M. (2004). The impact of extensive medial frontal lobe damage on 'Theory of Mind' and cognition. Brain *127*, 914-928.

Blair, R. J., Morris, J. S., Frith, C. D., Perrett, D. I., and Dolan, R. J. (1999). Dissociable neural responses to facial expressions of sadness and anger. Brain *122*, 883-893.

Blier, P., and de Montigny, C. (1994). Current advances and trends in the treatment of depression. Trends Pharmacol Sci *15*, 220-226.

Bloch, F., Hansen, W. W., and Packard, M. (1946). The nuclear induction experiment. Physical Review *70*, 474-485.

Bobes, M. A., Martin, M., Olivares, E., and Valdes-Sosa, M. (2000). Different scalp topography of brain potentials related to expression and identity matching of faces. Brain Res Cogn Brain Res *9*, 249-260.

Bodamer, J. (1947). Die Prosopagnosie. Arch Psychiatr Nervenkr *179*.

Bonda, E., Petrides, M., Ostry, D., and Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. J Neurosci *16*, 3737-3744.

Borelli, G. B. (1867). Annotazione alla precedente osservazione. Giornale d'Oftalmologia Italiano *10*, 112-117.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., Strauss, M. M., Hyman, S. E., and Rosen, B. R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. Neuron *17*, 875-887.

Brigham, J. C., and Barkowitz, P. (1978). Do "they all look alike?" - effect of race, sex, experience, and attitudes on the ability to recognize faces. J Appl Soc Psychol *8*, 306-318.

Brigham, J. C., and Malpass, R. S. (1985). The role of experience and contact in the recognition of faces of own- and other-race persons. J Social Issues *41*, 139–155.

Broca, P. (1878). Anatomie comparée: Le grand lobe limbique et la scissure limbique dans la série des mammiferes. Revue d'Anthropologie *1*, 385-408.

Broca, P. (1879). Sur les températures morbides locales. Bulletin de l'Académie de Medecine *8*, 1331-1347.

Broks, P., Young, A. W., Maratos, E. J., Coffey, P. J., Calder, A. J., Isaac, C. L., Mayes, A. R., Hodges, J. R., Montaldi, D., Cezayirli, E., *et al.* (1998). Face processing impairments after encephalitis: amygdala damage and recognition of fear. Neuropsychologia *36*, 59-70.

Brothers, L. (1989). A biological perspective on empathy. Am J Psychiatry *146*, 10-19.

Brothers, L. (1990). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. Concepts Neurosci *1*, 27-51.

Brothers, L., Ring, B., and Kling, A. (1990). Response of neurons in the macaque amygdala to complex social stimuli. Behav Brain Res *41*, 199-213.

Brothers, L., and Ring, B. (1993). Mesial temporal neurons in the macaque monkey with responses selective for aspects of social stimuli. Behav Brain Res *57*, 53-61.

Bruce, V., and Young, A. (1986). Understanding face recognition. Br J Psychol *77*, 305-327.

Brunet, E., Sarfati, Y., Hardy-Bayle, M. C., and Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. Neuroimage *11*, 157-166.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., and Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. Eur J Neurosci *13*, 400-404.

Buchel, C., Wise, R. J., Mummery, C. J., Poline, J. B., and Friston, K. J. (1996). Nonlinear regression in parametric activation studies. Neuroimage *4*, 60-66.

Buchel, C., Holmes, A. P., Rees, G., and Friston, K. J. (1998a). Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. Neuroimage *8*, 140-148.

Buchel, C., Morris, J., Dolan, R. J., and Friston, K. J. (1998b). Brain systems mediating aversive conditioning: an event-related fMRI study. Neuron *20*, 947-957.

Buchel, C., Dolan, R. J., Armony, J. L., and Friston, K. J. (1999). Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. J Neurosci *19*, 10869-10876.

Bush, G., Luu, P., and Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn Sci *4*, 215-222.

Buxton, R. B. (2002). Introduction to functional magnetic resonance imaging: Principles & techniques (Cambridge, UK, Cambridge University Press).

Calder, A., Young, A. W., Rowland, D., Perrett, D. I., Hodges, J. R., and Etcoff, N. L. (1996). Facial emotion recognition after bilateral amygdala damage - Differentially severe impairment of fear. Cognitive Neuropsychology *13*, 699-745.

Calder, A. J., Keane, J., Manes, F., Antoun, N., and Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. Nat Neurosci *3*, 1077-1078.

Calder, A. J., Lawrence, A. D., and Young, A. W. (2001). Neuropsychology of fear and loathing. Nat Rev Neurosci *2*, 352-363.

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., and Young, A. W. (2002). Reading the mind from eye gaze. Neuropsychologia *40*, 1129-1138.

Calder, A. J., Keane, J., Lawrence, A. D., and Manes, F. (2004). Impaired recognition of anger following damage to the ventral striatum. Brain *127*, 1958-1969.

Carey, D. P. (1996). 'Monkey see, monkey do' cells. Neurophysiology. Curr Biol *6*, 1087-1088.

Carmichael, S. T., and Price, J. L. (1995a). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. J Comp Neurol *363*, 615-641.

Carmichael, S. T., and Price, J. L. (1995b). Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. J Comp Neurol *363*, 642-664.

Casey, K. L. (1999). Forebrain mechanisms of nociception and pain: analysis through imaging. Proc Natl Acad Sci U S A *96*, 7668-7674.

Castelli, F., Happe, F., Frith, U., and Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage *12*, 314-325.

Castelli, F., Frith, C., Happe, F., and Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. Brain *125*, 1839-1849.

Chee, M. W., Soon, C. S., and Lee, H. L. (2003). Common and segregated neuronal networks for different languages revealed using functional magnetic resonance adaptation. J Cogn Neurosci *15*, 85-97.

Chen, N. K., Dickey, C. C., Yoo, S. S., Guttmann, C. R., and Panych, L. P. (2003). Selection of voxel size and slice orientation for fMRI in the presence of susceptibility field gradients: application to imaging of the amygdala. Neuroimage *19*, 817-825.

Chiroro, P., and Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. Q J Exp Psychol A: Hum Exp Psychol *48*, 879–894.

Cohen, L., Smith, M. J., and Leroux-Hugon, V. (2004). Paul Broca's thermometric crown. J Neurol Neurosurg Psychiatry *75*, 32.

Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. Cognition *31*, 187-276.

Cosmides, L., and Tooby, J. (1992). Cognitive adaptations for social exchange. In The Adapted Mind: Evolutionary Psychology and the Generation of Culture, J. H. Barchow, L. Cosmides, and J. Tooby, eds. (Oxford, Oxford University Press), pp. 163-228.

Cosmides, L., and Tooby, J. (2000). Evolutionary psychology and the emotions. In Handbook of Emotions, M. Lewis, and J. M. Haviland-Jones, eds. (NY, Guilford), pp. 91-115.

Craig, A. D. (2003). Interoception: the sense of the physiological condition of the body. Curr Opin Neurobiol *13*, 500-505.

Critchley, H. D., and Rolls, E. T. (1996a). Responses of primate taste cortex neurons to the astringent tastant tannic acid. Chem Senses *21*, 135-145.

Critchley, H. D., and Rolls, E. T. (1996b). Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. J Neurophysiol *75*, 1673-1686.

Critchley, H. D., and Rolls, E. T. (1996c). Olfactory neuronal responses in the primate orbitofrontal cortex: analysis in an olfactory discrimination task. J Neurophysiol *75*, 1659-1672.

Critchley, H., Daly, E., Phillips, M., Brammer, M., Bullmore, E., Williams, S., Van Amelsvoort, T., Robertson, D., David, A., and Murphy, D. (2000a). Explicit and implicit neural mechanisms for processing of social information from facial expressions: a functional magnetic resonance imaging study. Hum Brain Mapp *9*, 93-105.

Critchley, H. D., Corfield, D. R., Chandler, M. P., Mathias, C. J., and Dolan, R. J. (2000b). Cerebral correlates of autonomic cardiovascular arousal: a functional neuroimaging investigation in humans. J Physiol *523*, 259-270.

Critchley, H. D., Elliott, R., Mathias, C. J., and Dolan, R. J. (2000c). Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. J Neurosci *20*, 3033-3040.

Critchley, H. D., Mathias, C. J., and Dolan, R. J. (2001a). Neural activity in the human brain relating to uncertainty and arousal during anticipation. Neuron *29*, 537-545.

Critchley, H. D., Mathias, C. J., and Dolan, R. J. (2001b). Neuroanatomical basis for first- and second-order representations of bodily states. Nat Neurosci *4*, 207-212.

Critchley, H. D., Mathias, C. J., and Dolan, R. J. (2002). Fear conditioning in humans: the influence of awareness and autonomic arousal on functional neuroanatomy. Neuron *33*, 653-663.

Critchley, H. D., Mathias, C. J., Josephs, O., O'Doherty, J., Zanini, S., Dewar, B. K., Cipolotti, L., Shallice, T., and Dolan, R. J. (2003). Human cingulate cortex and autonomic control: converging neuroimaging and clinical evidence. Brain *126*, 2139-2152.

Critchley, H. (2004). The human cortex responds to an interoceptive challenge. Proc Natl Acad Sci U S A *101*, 6333-6334.

Critchley, H. D., Rotshtein, P., Nagai, Y., O'Doherty, J., Mathias, C. J., and Dolan, R. J. (2005). Activity in the human brain predicting differential heart rate responses to emotional facial expressions. Neuroimage *24*, 751-762.

Cunningham, W. A., Johnson, M. K., Gatenby, J. C., Gore, J. C., and Banaji, M. R. (2003). Neural components of social evaluation. J Pers Soc Psychol *85*, 639-649.

Curtis, V., Aunger, R., and Rabie, T. (2004). Evidence that disgust evolved to protect from risk of disease. Biol Lett *271*, 131-133.

Damasio, A. R., Tranel, D., and Damasio, H. (1990a). Face agnosia and the neural substrates of memory. Annu Rev Neurosci *13*, 89-109.

Damasio, A. R., Tranel, D., and Damasio, H. (1990b). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. Behav Brain Res *41*, 81-94.

Damasio, A. R. (1994). Descartes' Error: Emotion, Reason and the Human Brain (New York, G.P. Putnam's Sons).

Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., and Damasio, A. R. (1994). The return of Phineas Gage: clues about the brain from the skull of a famous patient. Science *264*, 1102-1105.

Damasio, A. (1999). The Feeling of What Happens: Body and Emotion in the Making of Consciousness (New York, Harcourt Brace).

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J., and Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self- generated emotions. Nat Neurosci *3*, 1049-1056.

Damasio, A. R. (2001). Fundamental feelings. Nature *413*, 781.

Darwin, C. R. (1872). The Expression of the Emotions in Man and Animals (London, John Murray).

Davis, M., and Whalen, P. J. (2001). The amygdala: vigilance and emotion. Mol Psychiatry *6*, 13-34.

de Gelder, B., Frissen, I., Barton, J., and Hadjikhani, N. (2003). A modulatory role for facial expressions in prosopagnosia. Proc Natl Acad Sci U S A *100*, 13105-13110.

De Renzi, E., Faglioni, P., Grossi, D., and Nichelli, P. (1991). Apperceptive and associative forms of prosopagnosia. Cortex *27*, 213-221.

Decety, J., and Chaminade, T. (2003). Neural correlates of feeling sympathy. Neuropsychologia *41*, 127-138.

Deichmann, R., Gottfried, J. A., Hutton, C., and Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage *19*, 430-441.

Dennett, D. C. (1989). The Intentional Stance (Boston, MIT Press).

Deruelle, C., Rondan, C., Gepner, B., and Tardif, C. (2004). Spatial frequency and face processing in children with autism and Asperger syndrome. J Autism Dev Disord *34*, 199-210.

Desimone, R., Albright, T. D., Gross, C. G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci *4*, 2051-2062.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. Proc Natl Acad Sci U S A *93*, 13494-13499.

Dicks, D., Myers, R. E., and Kling, A. (1968). Uncus and amygdala lesions: effects on social behavior in the free-ranging rhesus monkey. Science *165*, 69-71.

Dimitrov, M., Phipps, M., Zahn, T. P., and Grafman, J. (1999). A Thoroughly Modern Gage. Neurocase *5*, 345-353.

Dodds, C., Machado, L., Rafal, R., and Ro, T. (2002). A temporal/nasal asymmetry for blindsight in a localisation task: evidence for extrageniculate mediation. Neuroreport *13*, 655-658.

Dolan, R. J., Fletcher, P., Morris, J., Kapur, N., Deakin, J. F., and Frith, C. D. (1996). Neural activation during covert processing of positive emotional facial expressions. *4*, 194-200.

Dolan, R. J., Lane, R., Chua, P., and Fletcher, P. (2000). Dissociable temporal lobe activations during emotional episodic memory retrieval. Neuroimage *11*, 203-209.

Dolan, R. J., Morris, J. S., and de Gelder, B. (2001). Crossmodal binding of fear in voice and face. Proc Natl Acad Sci U S A *98*, 10006-10010.

Dolan, R. J. (2002). Emotion, cognition, and behavior. Science *298*, 1191-1194.

Dolan, R. J., and Vuilleumier, P. (2003). Amygdala automaticity in emotional processing. Ann N Y Acad Sci *985*, 348-355.

Donaldson, D. I. (2004). Parsing brain activity with fMRI and mixed designs: what kind of a state is neuroimaging in? Trends Neurosci *27*, 442-444.

Donders, F. C. (1969). On the speed of mental processes. Acta Psychol (Amst) *30*, 412-431.

Drevets, W. C., Price, J. L., Simpson, J. R., Jr., Todd, R. D., Reich, T., Vannier, M., and Raichle, M. E. (1997). Subgenual prefrontal cortex abnormalities in mood disorders. Nature *386*, 824-827.

Drevets, W. C., Bogers, W., and Raichle, M. E. (2002). Functional anatomical correlates of antidepressant drug treatment assessed using PET measures of regional glucose metabolism. Eur Neuropsychopharmacol *12*, 527-544.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. Nat Rev Neurosci *2*, 820-829.

Duvernoy, H. M. (1999). The Human Brain (Vienna, Springer-Verlag).

Eger, E., Schyns, P. G., and Kleinschmidt, A. (2004). Scale invariant adaptation in fusiform face-responsive regions. Neuroimage *22*, 232-242.

Eifuku, S., De Souza, W. C., Tamura, R., Nishijo, H., and Ono, T. (2004). Neuronal correlates of face identification in the monkey anterior temporal cortical areas. J Neurophysiol *91*, 358-371.

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. Clin Neurophysiol *111*, 694-705.

Eimer, M., and Holmes, A. (2002). An ERP study on the time course of emotional face processing. Neuroreport *13*, 427-431.

Eimer, M., Holmes, A., and McGlone, F. P. (2003). The role of spatial attention in the processing of facial expression: an ERP study of rapid brain responses to six basic emotions. Cogn Affect Behav Neurosci *3*, 97-110.

Ekman, P., and Friesen, W. V. (1975). Pictures of facial affect (Palo Alto, CA, Consulting Psychologists Press).

Ekman, P. (1993). Facial expression and emotion. Am Psychol *48*, 384-392.

Elliott, R., and Dolan, R. J. (1998). Neural response during preference and memory judgments for subliminally presented stimuli: a functional neuroimaging study. J Neurosci *18*, 4697-4704.

Elliott, R., Newman, J. L., Longe, O. A., and Deakin, J. F. (2003). Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: a parametric functional magnetic resonance imaging study. J Neurosci *23*, 303-307.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. Neurosci Biobehav Rev *24*, 581-604.

Epstein, R., Graham, K. S., and Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. Neuron *37*, 865-876.

Eslinger, P. J., and Damasio, A. R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. Neurology *35*, 1731-1741.

Eslinger, P. J. (1998). Neurological and neuropsychological bases of empathy. Eur Neurol *39*, 193-199.

Esteves, F., and Ohman, A. (1993). Masking the face: recognition of emotional facial expressions as a function of the parameters of backward masking. ScandJPsychol *34*, 1-18.

Esteves, F., Parra, C., Dimberg, U., and Ohman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. Psychophysiology *31*, 375-385.

Etcoff, N. L. (1984). Selective attention to facial identity and facial emotion. Neuropsychologia *22*, 281-295.

Etcoff, N. L., and Magee, J. J. (1992). Categorical perception of facial expressions. Cognition *44*, 227-240.

Fallon, J. H., and Moore, R. Y. (1978). Catecholamine innervation of the basal forebrain. IV. Topography of the dopamine projection to the basal forebrain and neostriatum. J Comp Neurol *180*, 545-580.

Farah, M. J. (1990). Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision (Cambridge, Mass., MIT Press).

Fine, C., and Blair, R. J. (2000). The cognitive and emotional effects of amygdala damage. Neurocase *6*, 435-438.

Fine, C., Lumsden, J., and Blair, R. J. (2001). Dissociation between 'theory of mind' and executive functions in a patient with early left amygdala damage. Brain *124*, 287-298.

Fink, B., and Penton-Voak, I. (2002). Evolutionary psychology of facial attractiveness. Curr Dir Psychol Sci *11*, 154-158.

Fink, G. R., Markowitsch, H. J., Reinkemeier, M., Bruckbauer, T., Kessler, J., and Heiss, W. D. (1996). Cerebral representation of one's own past: neural networks involved in autobiographical memory. J Neurosci *16*, 4275-4282.

Fisher, R. A. (1966). The design of experiments, 8th edn (Edinburgh, Oliver & Boyd).

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., and Frith, C. D. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition *57*, 109-128.

Frahm, J., Bruhn, H., Merboldt, K. D., and Hanicke, W. (1992). Dynamic MR imaging of human brain oxygenation during rest and photic stimulation. J Magn Reson Imaging *2*, 501-505.

Francis, S., Rolls, E. T., Bowtell, R., McGlone, F., O'Doherty, J., Browning, A., Clare, S., and Smith, E. (1999). The representation of pleasant touch in the brain and its relationship with taste and olfactory areas. Neuroreport *10*, 453-459.

Fredrikson, M., Furmark, T., Olsson, M. T., Fischer, H., Andersson, J., and Langstrom, B. (1998). Functional neuroanatomical correlates of electrodermal activity: a positron emission tomographic study. Psychophysiology *35*, 179-185.

Freud, S. (1920). Beyond the pleasure principle (London, The Hogarth Press and the Institute of Psychoanalysis).

Fried, I., MacDonald, K. A., and Wilson, C. L. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. Neuron *18*, 753-765.

Frieze, I. H., lson, J. E., and Russell, J. (1991). Attractiveness and income for men and women in management. J Appl Soc Psychol *21*, 1039-1057.

Friston, K., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D., and Frackowiak, R. S. J. (1995a). Spatial registration and normalization of images. Hum Brain Mapp *2*, 165-189.

Friston, K., Holmes, A. P., Worsley, K., Poline, J.-B., Frith, C., and Frackowiak, R. S. J. (1995b). Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp *2*, 189-210.

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S., and Dolan, R. J. (1996). The trouble with cognitive subtraction. Neuroimage *4*, 97-104.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., and Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. Neuroimage *6*, 218-229.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. (1998a). Event-related fMRI: characterizing differential responses. Neuroimage *7*, 30-40.

Friston, K. J., Josephs, O., Rees, G., and Turner, R. (1998b). Nonlinear event-related responses in fMRI. Magn Reson Med *39*, 41-52.

Friston, K. J., Holmes, A. P., and Worsley, K. J. (1999a). How many subjects constitute a study? Neuroimage *10*, 1-5.

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N., and Dale, A. M. (1999b). Stochastic designs in event-related fMRI. Neuroimage *10*, 607-619.

Frith, C. D., and Frith, U. (1999). Interacting minds--a biological basis. Science *286*, 1692-1695.

Fuller, R. W. (1995). Serotonin uptake inhibitors: uses in clinical therapy and in laboratory research. Prog Drug Res *45*, 167-204.

Fulton, J. F. (1928). Observations upon the vascularity of the human occipital lobe during visual activity. Brain *51*, 310-320.

Funahashi, S. (2001). Neuronal mechanisms of executive control by the prefrontal cortex. Neurosci Res *39*, 147-165.

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., and Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. Neuropsychologia *38*, 11-21.

Gallagher, H. L., Jack, A. I., Roepstorff, A., and Frith, C. D. (2002). Imaging the intentional stance in a competitive game. Neuroimage *16*, 814-821.

Gallagher, H. L., and Frith, C. D. (2003). Functional imaging of 'theory of mind'. Trends Cogn Sci *7*, 77-83.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. Brain *119*, 593-609.

Gallese, V. (2003). The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity. Psychopathology *36*, 171-180.

Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. Trends Cogn Sci *8*, 396-403.

Ganel, T., and Goshen-Gottstein, Y. (2004). Effects of familiarity on the perceptual integrality of the identity and expression of faces: the parallel-route hypothesis revisited. J Exp Psychol Hum Percept Perform *30*, 583-597.

Garavan, H., Pendergrass, J. C., Ross, T. J., Stein, E. A., and Risinger, R. C. (2001). Amygdala response to both positively and negatively valenced stimuli. Neuroreport *12*, 2779-2783.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat Neurosci *2*, 568-573.

Gauthier, I., Skudlarski, P., Gore, J. C., and Anderson, A. W. (2000a). Expertise for cars and birds recruits brain areas involved in face recognition. Nat Neurosci *3*, 191-197.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., and Anderson, A. W. (2000b). The fusiform "face area" is part of a network that processes faces at the individual level. J Cogn Neurosci *12*, 495-504.

Gazzaniga, M. S., Ivry, R. B., and Mangun, G. R. (1998). Cognitive Neuroscience: The biology of the mind (New York, W.W. Norton & Company).

George, M. S., Ketter, T. A., Parekh, P. I., Horwitz, B., Herscovitch, P., and Post, R. M. (1995). Brain activity during transient sadness and happiness in healthy women. Am J Psychiatry *152*, 341-351.

George, N., Evans, J., Fiori, N., Davidoff, J., and Renault, B. (1996). Brain events related to normal and moderately scrambled faces. Brain Res Cogn Brain Res *4*, 65-76.

George, N., Dolan, R. J., Fink, G. R., Baylis, G. C., Russell, C., and Driver, J. (1999). Contrast polarity and face recognition in the human fusiform gyrus. Nat Neurosci *2*, 574-580.

George, N., Driver, J., and Dolan, R. J. (2001). Seen gaze-direction modulates fusiform activity and its coupling with other brain areas during face processing. Neuroimage *13*, 1102-1112.

Glaser, D., and Friston, K. (2004). Variance Components. In Human Brain Function (second edition), R. S. J. Frackowiak, K. Friston, C. Frith, R. Dolan, C. Price, S. Zeki, J. Ashburner, and W. Penney, eds.

Goel, V., Grafman, J., Sadato, N., and Hallett, M. (1995). Modeling other minds. Neuroreport *6*, 1741-1746.

Golby, A. J., Gabrieli, J. D., Chiao, J. Y., and Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. Nat Neurosci *4*, 845-850.

Goren, C. C., Sarty, M., and Wu, P. Y. K. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. Pediatrics *56*, 544-549.

Gorno-Tempini, M. L., Pradelli, S., Serafini, M., Pagnoni, G., Baraldi, P., Porro, C., Nicoletti, R., Umita, C., and Nichelli, P. (2001). Explicit and incidental facial expression processing: an fMRI study. Neuroimage *14*, 465-473.

Gottfried, J. A., Deichmann, R., Winston, J. S., and Dolan, R. J. (2002). Functional heterogeneity in human olfactory cortex: an event-related functional magnetic resonance imaging study. J Neurosci *22*, 10819-10828.

Gottfried, J. A., O'Doherty, J., and Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science *301*, 1104-1107.

Grafton, S. T., Arbib, M. A., Fadiga, L., and Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. Exp Brain Res *112*, 103-111.

Grattan, L. M., and Eslinger, P. J. (1992). Long-term psychological consequences of childhood frontal lobe lesion in patient DT. Brain Cogn *20*, 185-195.

Grelotti, D. J., Gauthier, I., and Schultz, R. T. (2002). Social interest and the development of cortical face specialization: what autism teaches us about face processing. Dev Psychobiol *40*, 213-225.

Grill-Spector, K., and Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol (Amst) *107*, 293-321.

Grill-Spector, K. (2003). The neural basis of object perception. Curr Opin Neurobiol *13*, 159-166.

Gross, C. G., Bender, D. B., and Rocha-Miranda, C. E. (1969). Visual Receptive Fields of Neurons in Inferotemporal Cortex of the Monkey. Science *166*, 1303-1306.

Gross, C. G., Rocha-Miranda, C. E., and Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. J Neurophysiol *35*, 96-111.

Gur, R. C., Schroeder, L., Turner, T., McGrath, C., Chan, R. M., Turetsky, B. I., Alsop, D., Maldjian, J., and Gur, R. E. (2002). Brain activation during facial emotion processing. Neuroimage *16*, 651-662.

Gusnard, D. A., Akbudak, E., Shulman, G. L., and Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. Proc Natl Acad Sci U S A *98*, 4259-4264.

Hadjikhani, N., and de Gelder, B. (2003). Seeing fearful body expressions activates the fusiform cortex and amygdala. Curr Biol *13*, 2201-2205.

Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B., Marinkovic, K., and Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. Hum Brain Mapp *7*, 29-37.

Hamann, S. B., Stefanacci, L., Squire, L. R., Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. (1996). Recognizing facial emotion. Nature *379*, 497.

Hamann, S. B., Ely, T. D., Grafton, S. T., and Kilts, C. D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. Nat Neurosci *2*, 289-293.

Hamann, S. B., Ely, T. D., Hoffman, J. M., and Kilts, C. D. (2002). Ecstasy and agony: activation of the human amygdala in positive and negative emotion. Psychol Sci *13*, 135-141.

Handwerker, D. A., Ollinger, J. M., and D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage *21*, 1639-1651.

Happe, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., Dolan, R., Frackowiak, R., and Frith, C. (1996). 'Theory of mind' in the brain. Evidence from a PET scan study of Asperger syndrome. Neuroreport *8*, 197-201.

Hariri, A. R., Bookheimer, S. Y., and Mazziotta, J. C. (2000). Modulating emotional responses: effects of a neocortical network on the limbic system. Neuroreport *11*, 43-48.

Hariri, A. R., Mattay, V. S., Tessitore, A., Fera, F., and Weinberger, D. R. (2003). Neocortical modulation of the amygdala response to fearful stimuli. Biol Psychiatry *53*, 494-501.

Harlow, J. M. (1848). Passage of an iron rod through the head. Boston Medical Surgery J *39*, 389-393.

Harlow, J. M. (1868). Recovery after severe injury to the head (Boston, Massachusetts Medical Society).

Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., and Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. Neuroreport *11*, 2351-2355.

Hasselmo, M. E., Rolls, E. T., and Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. Behav Brain Res *32*, 203-218.

Hatfield, T., Han, J. S., Conley, M., Gallagher, M., and Holland, P. (1996). Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. J Neurosci *16*, 5256-5265.

Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., Herscovitch, P., Schapiro, M. B., and Rapoport, S. I. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. Proc Natl Acad Sci U S A *88*, 1621-1625.

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., and Grady, C. L. (1994). The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. J Neurosci *14*, 6336-6353.

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., and Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. Neuron *22*, 189-199.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. Trends Cogn Sci *4*, 223-233.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. Biol Psychiatry *51*, 59-67.

Heims, H. C., Critchley, H. D., Dolan, R., Mathias, C. J., and Cipolotti, L. (2004). Social and motivational functioning is not critically dependent on feedback of autonomic responses: neuropsychological evidence from patients with pure autonomic failure. Neuropsychologia *42*, 1979-1988.

Henderson, J. A., and Anglin, J. M. (2003). Facial attractiveness predicts longevity. Evol Hum Behav *24*, 351-356.

Henson, R., Shallice, T., and Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. Science *287*, 1269-1272.

Henson, R. N., Goshen-Gottstein, Y., Ganel, T., Otten, L. J., Quayle, A., and Rugg, M. D. (2003). Electrophysiological and haemodynamic correlates of face perception, recognition and priming. Cereb Cortex *13*, 793-805.

Henson, R. N., and Rugg, M. D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. Neuropsychologia *41*, 263-270.

Henson, R. N. A. (2003). Neuroimaging studies of priming. Prog Neurobiol *70*, 53-81.

Heywood, C. A., and Cowey, A. (1992). The role of the 'face-cell' area in the discrimination and recognition of faces by monkeys. Philos Trans R Soc Lond B Biol Sci *335*, 31-37.

Hof, P. R., Mufson, E. J., and Morrison, J. H. (1995). Human orbitofrontal cortex: cytoarchitecture and quantitative immunohistochemical parcellation. J Comp Neurol *359*, 48-68.

Hoffman, E. A., and Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat Neurosci *3*, 80-84.

Holmes, A., and Friston, K. (1998). Generalisability, Random Effects & Population Inference. Neuroimage *7*, S754.

Holmes, A., Vuilleumier, P., and Eimer, M. (2003). The processing of emotional facial expression is gated by spatial attention: evidence from event-related brain potentials. Brain Res Cogn Brain Res *16*, 174-184.

Hornak, J., Rolls, E. T., and Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. Neuropsychologia *34*, 247-261.

Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O'Doherty, J., Bullock, P. R., and Polkey, C. E. (2003). Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. Brain *126*, 1691-1712.

Hornak, J. P. (2003). The Basics of MRI, pp. http://www.cis.rit.edu/htbooks/mri/.

Howard, M. A., Cowell, P. E., Boucher, J., Broks, P., Mayes, A., Farrant, A., and Roberts, N. (2000). Convergent neuroanatomical and behavioural evidence of an amygdala hypothesis of autism. Neuroreport *11*, 2931-2935.

Hu, X., Le, T. H., and Ugurbil, K. (1997). Evaluation of the early response in fMRI in individual subjects using short stimulus duration. Magn Reson Med *37*, 877-884.

Hughlings Jackson, J. (1872). Case of disease of the brain – left hemiplegia – mental affection. Medical Times and Gazette *May 4th*, 513-514.

Humphrey, N. (1983). Consciousness regained: Chapters in the development of mind (New York, Oxford University Press).

Iidaka, T., Omori, M., Murata, T., Kosaka, H., Yonekura, Y., Okada, T., and Sadato, N. (2001). Neural interaction of the amygdala with the prefrontal and temporal cortices in the processing of facial expressions as revealed by fMRI. J Cogn Neurosci *13*, 1035-1047.

Isenberg, N., Silbersweig, D., Engelien, A., Emmerich, S., Malavade, K., Beattie, B., Leon, A. C., and Stern, E. (1999). Linguistic threat activates the human amygdala. Proc Natl Acad Sci U S A *96*, 10456-10459.

Itier, R. J., and Taylor, M. J. (2004). Source analysis of the N170 to faces and objects. Neuroreport *15*, 1261-1265.

James, W. (1884). What is an Emotion? Mind *9*, 188-205.

Jellema, T., Baker, C. I., Wicker, B., and Perrett, D. I. (2000). Neural representation for the perception of the intentionality of actions. Brain Cogn *44*, 280-302.

Jellema, T., and Perrett, D. I. (2003). Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? Neuropsychologia *41*, 1728-1737.

Jezzard, P., and Clare, S. (2002). Principles of nuclear magnetic resonance and MRI. In Functional MRI: an introduction to methods, P. Jezzard, P. M. Matthews, and S. M. Smith, eds. (Oxford, Oxford University Press), pp. 67-92.

Johnson, M. H., Dziurawiec, S., Ellis, H., and Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. Cognition *40*, 1-19.

Johnson, M. H., and Morton, J. (1991). Biology and Cognitive Development: The Case of Face Recognition (London, Basil Blackwell Ltd).

Johnson, M. K., Nolde, S. F., Mather, M., Kounios, J., Schacter, D. L., and Curran, T. (1997). Test format can affect the similarity of brain activity associated with true and false recognition memory. Psychol Sci *8*, 250-257.

Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., and Prigatano, G. P. (2002). Neural correlates of self-reflection. Brain *125*, 1808-1814.

Johnsrude, I. S., Owen, A. M., White, N. M., Zhao, W. V., and Bohbot, V. (2000). Impaired preference conditioning after anterior temporal lobe resection in humans. J Neurosci *20*, 2649-2656.

Joseph, J. E. (2001). Functional neuroimaging studies of category specificity in object recognition: a critical review and meta-analysis. Cogn Affect Behav Neurosci *1*, 119-136.

Josephs, O., Turner, R., and Friston, K. (1997). Event-related fMRI. Hum Brain Mapp *5*, 243-248.

Josephs, O., and Henson, R. N. (1999). Event-related functional magnetic resonance imaging: modelling, inference and optimization. Philos Trans R Soc Lond B Biol Sci *354*, 1215-1228.

Kampe, K. K., Frith, C. D., Dolan, R. J., and Frith, U. (2001). Reward value of attractiveness and gaze. Nature *413*, 589.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci *17*, 4302-4311.

Kanwisher, N. (2000). Domain specificity in face perception. Nat Neurosci *3*, 759-763.

Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Ito, K., Fukuda, H., Kojima, S., and Nakamura, K. (1999). The human amygdala plays an important role in gaze monitoring. A PET study. Brain *122*, 779-783.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., and Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. J Cogn Neurosci *14*, 785-794.

Kendrick, K. M., and Baldwin, B. A. (1987). Cells in temporal cortex of conscious sheep can respond preferentially to the sight of faces. Science *236*, 448-450.

Kesler-West, M. L., Andersen, A. H., Smith, C. D., Avison, M. J., Davis, C. E., Kryscio, R. J., and Blonder, L. X. (2001). Neural substrates of facial emotion processing using fMRI. Brain Res Cogn Brain Res *11*, 213-226.

Keysers, C., and Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. Trends Cogn Sci *8*, 501-507.

Kiebel, S. J., Poline, J. B., Friston, K. J., Holmes, A. P., and Worsley, K. J. (1999). Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. Neuroimage *10*, 756-766.

Kiebel, S. J., Glaser, D. E., and Friston, K. J. (2003). A heuristic for the degrees of freedom of statistics based on multiple variance parameters. Neuroimage *20*, 591-600.

Kim, D. S., Duong, T. Q., and Kim, S. G. (2000). High-resolution mapping of iso-orientation columns by fMRI. Nat Neurosci *3*, 164-169.

Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., and Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. Neuroreport *14*, 2317-2322.

Klüver, H., and Bucy, P. C. (1939). Preliminary analysis of the temporal lobes in monkeys. Arch Neurol Psychiatry *42*, 979-1000.

Kolb, B., Milner, B., and Taylor, L. (1983). Perception of faces by patients with localized cortical excisions. Can J Psychol *37*, 8-18.

Kourtzi, Z., and Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. Science *293*, 1506-1509.

Kringelbach, M. L., and Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. Prog Neurobiol *72*, 341-372.

Krolak-Salmon, P., Henaff, M. A., Vighetto, A., Bertrand, O., and Mauguiere, F. (2004). Early Amygdala Reaction to Fear Spreading in Occipital, Temporal, and Frontal Cortex; A Depth Electrode ERP Study in Human. Neuron *42*, 665-676.

Kucharska-Pietura, K., Phillips, M. L., Gernand, W., and David, A. S. (2003). Perception of emotions from faces and voices following unilateral brain damage. Neuropsychologia *41*, 1082-1090.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., and et al. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci U S A *89*, 5675-5679.

Lane, R. D., Fink, G. R., Chau, P. M., and Dolan, R. J. (1997a). Neural activation during selective attention to subjective emotional responses. Neuroreport *8*, 3969-3972.

Lane, R. D., Reiman, E. M., Ahern, G. L., Schwartz, G. E., and Davidson, R. J. (1997b). Neuroanatomical correlates of happiness, sadness, and disgust. Am J Psychiatry *154*, 926-933.

Lane, R. D., Chua, P. M., and Dolan, R. J. (1999). Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. Neuropsychologia *37*, 989-997.

Lauterbur, P. C. (1973). Image formation by induced local interactions: examples employing nuclear magnetic resonance. Nature *242*, 190-191.

LeDoux, J. E., Sakaguchi, A., and Reis, D. J. (1984). Subcortical efferent projections of the medial geniculate nucleus mediate emotional responses conditioned to acoustic stimuli. J Neurosci *4*, 683-698.

LeDoux, J. E. (1996). The Emotional Brain (New York, Schuster and Schuster).

LeDoux, J. E. (2000). Emotion circuits in the brain. Annu Rev Neurosci *23*, 155-184.

Leibenluft, E., Gobbini, M. I., Harrison, T., and Haxby, J. V. (2004). Mothers' neural activation in response to pictures of their children and other children. Biol Psychiatry *56*, 225-232.

Leonard, C. M., Rolls, E. T., Wilson, F. A., and Baylis, G. C. (1985). Neurons in the amygdala of the monkey with responses selective for faces. Behav Brain Res *15*, 159-176.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature *412*, 150-157.

Logothetis, N. K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. Philos Trans R Soc Lond B Biol Sci *357*, 1003-1037.

Lueck, C. J., Zeki, S., Friston, K. J., Deiber, M. P., Cope, P., Cunningham, V. J., Lammertsma, A. A., Kennard, C., and Frackowiak, R. S. (1989). The colour centre in the cerebral cortex of man. Nature *340*, 386-389.

Lundqvist, D., and Litton, J.-E. (1998). The Averaged Karolinska Directed Emotional Faces (Department of Clinical Neuroscience, Psychology Section, Karolinska Institute).

MacClean, P. D. (1949). Psychosomatic disease and the "visceral brain": recent developments bearing on the Papez theory of emotion. Psychosom Med *11*, 338-353.

MacClean, P. D. (1970). The triune brain, emotion and scientific bias. In Neurosciences: Second Study Program, F. O. Schmitt, ed. (New York, Rockefeller University Press), pp. 336-349.

Maddock, R. J. (1999). The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. Trends Neurosci *22*, 310-316.

Magistretti, P. J., and Pellerin, L. (1999). Astrocytes Couple Synaptic Activity to Glucose Utilization in the Brain. News Physiol Sci *14*, 177-182.

Maguire, E. A. (2001a). Neuroimaging studies of autobiographical event memory. Philos Trans R Soc Lond B Biol Sci *356*, 1441-1451.

Maguire, E. A. (2001b). The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. Scand J Psychol *42*, 225-238.

Malonek, D., and Grinvald, A. (1996). Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: implications for functional brain mapping. Science *272*, 551-554.

Malpass, R. S., and Kravitz, J. (1969). Recognition for faces of own and other race. J Pers Soc Psychol *13*, 330-334.

Mansfield, P., and Grannell, P. K. (1973). NMR 'diffraction' in solids? Journal of Physics C Solid State *6*, L422-L426.

Mansfield, P. (1977). Multi-planar image formation using NMR spin echoes. J Physical Chem *10*, L55-L58.

Martin, A., and Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. Cogn Neuropsychol *20*, 575-587.

McCabe, K., Houser, D., Ryan, L., Smith, V., and Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. Proc Natl Acad Sci U S A *98*, 11832-11835.

McCarthy, G., Puce, A., Belger, A., and Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. Cereb Cortex *9*, 431-444.

Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., and Buckner, R. L. (2000). Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. Neuroimage *11*, 735-759.

Mirenowicz, J., and Schultz, W. (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. Nature *379*, 449-451.

Moll, J., Eslinger, P. J., and Oliveira-Souza, R. (2001). Frontopolar and anterior temporal cortex activation in a moral judgment task: preliminary functional MRI results in normal subjects. Arq Neuropsiquiatr *59*, 657-664.

Moll, J., de Oliveira-Souza, R., Bramati, I. E., and Grafman, J. (2002a). Functional networks in emotional moral and nonmoral social judgments. Neuroimage *16*, 696-703.

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A., and Pessoa, L. (2002b). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. J Neurosci *22*, 2730-2736.

Morecraft, R. J., Geula, C., and Mesulam, M. M. (1992). Cytoarchitecture and neural afferents of orbitofrontal cortex in the brain of the monkey. J Comp Neurol *323*, 341-358.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., and Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. Nature *383*, 812-815.

Morris, J. S., Friston, K. J., Buchel, C., Frith, C. D., Young, A. W., Calder, A. J., and Dolan, R. J. (1998a). A neuromodulatory role for the human amygdala in processing emotional facial expressions. Brain *121*, 47-57.

Morris, J. S., Ohman, A., and Dolan, R. J. (1998b). Conscious and unconscious emotional learning in the human amygdala. Nature *393*, 467-470.

Morris, J. S., Ohman, A., and Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. Proc Natl Acad Sci U S A *96*, 1680-1685.

Morris, J. S., Buchel, C., and Dolan, R. J. (2001a). Parallel neural responses in amygdala subregions and sensory cortex during implicit fear conditioning. Neuroimage *13*, 1044-1052.

Morris, J. S., DeGelder, B., Weiskrantz, L., and Dolan, R. J. (2001b). Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. Brain *124*, 1241-1252.

Morrison, D. J., and Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. Psychon Bull Rev *8*, 454-469.

Mosso, A. (1881). Über den Kreislauf des Blutes in Menschlichen Gehirn (Leipzig, Verlag von Viet).

Mouchetant-Rostaing, Y., and Giard, M. H. (2003). Electrophysiological correlates of age and gender perception on human faces. J Cogn Neurosci *15*, 900-910.

Munte, T. F., Brack, M., Grootheer, O., Wieringa, B. M., Matzke, M., and Johannes, S. (1998). Brain potentials reveal the timing of face identity and expression judgments. Neurosci Res *30*, 25-34.

Naccache, L., and Dehaene, S. (2001). The priming method: imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. Cereb Cortex *11*, 966-974.

Nakamura, K., Mikami, A., and Kubota, K. (1992). Activity of single neurons in the monkey amygdala during performance of a visual discrimination task. J Neurophysiol *67*, 1447-1463.

Nakamura, K., Kawashima, R., Nagumo, S., Ito, K., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Kubota, K., Fukuda, H., and Kojima, S. (1998). Neuroanatomical correlates of the assessment of facial attractiveness. Neuroreport *9*, 753-757.

Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., Hatano, K., Ito, K., Fukuda, H., Schormann, T., and Zilles, K. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing. A PET study. Brain *123*, 1903-1912.

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Nagumo, S., Kubota, K., Fukuda, H., Ito, K., and Kojima, S. (2001). Neural substrates for recognition of familiar voices: a PET study. Neuropsychologia *39*, 1047-1054.

Narumoto, J., Okada, T., Sadato, N., Fukui, K., and Yonekura, Y. (2001). Attention to emotion modulates fMRI activity in human right superior temporal sulcus. Brain Res Cogn Brain Res *12*, 225-231.

Nitschke, J. B., Nelson, E. E., Rusch, B. D., Fox, A. S., Oakes, T. R., and Davidson, R. J. (2004). Orbitofrontal cortex tracks positive mood in mothers viewing pictures of their newborn infants. Neuroimage *21*, 583-592.

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., and Andrews, C. (2001a). Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci *4*, 95-102.

O'Doherty, J., Rolls, E. T., Francis, S., Bowtell, R., and McGlone, F. (2001b). Representation of pleasant and aversive taste in the human brain. J Neurophysiol *85*, 1315-1321.

O'Doherty, J., Critchley, H., Deichmann, R., and Dolan, R. J. (2003a). Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. J Neurosci *23*, 7931-7939.

O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M., and Dolan, R. J. (2003b). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. Neuropsychologia *41*, 147-155.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science *304*, 452-454.

O'Doherty, J. P., Deichmann, R., Critchley, H. D., and Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. Neuron *33*, 815-826.

Ochsner, K. N., Bunge, S. A., Gross, J. J., and Gabrieli, J. D. (2002). Rethinking feelings: an FMRI study of the cognitive regulation of emotion. J Cogn Neurosci *14*, 1215-1229.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci U S A *87*, 9868-9872.

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., and Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci U S A *89*, 5951-5955.

Ohman, A. (2002). Automaticity and the amygdala: Nonconscious responses to emotional faces. Curr Dir Psychol Sci *11*, 62-66.

Ojemann, J. G., Ojemann, G. A., and Lettich, E. (1992). Neuronal activity related to faces and matching in human right nondominant temporal cortex. Brain *115*, 1-13.

Ono, T., and Nishijo, H. (2000). Neurophysiological basis of emotion in primates: Neuronal responses in the monkey amygdala and anterior cingulate cortex. In The New Cognitive Neurosciences, M. S. Gazzaniga, ed. (Cambridge, MA, MIT Press), pp. 1099-1114.

Palmeri, T. J., and Gauthier, I. (2004). Visual object understanding. Nat Rev Neurosci *5*, 291-303.

Panksepp, J. (2004). Basic affects and the instinctual emotional systems of the brain: The primordial sources of sadness, joy, and seeking. In Feelings and Emotions: The Amsterdam Symposium, A. S. R. Manstead, N. Fridja, and A. Fischer, eds. (Cambridge, Cambridge University Press), pp. 174-193.

Papez, J. W. (1937). A proposed mechanism of emotion. Arch Neurol Psychiatry *38*, 725-743.

Parkinson, J. A., Crofts, H. S., McGuigan, M., Tomic, D. L., Everitt, B. J., and Roberts, A. C. (2001). The role of the primate amygdala in conditioned reinforcement. J Neurosci *21*, 7770-7780.

Pasley, B. N., Mayes, L. C., and Schultz, R. T. (2004). Subcortical discrimination of unperceived objects during binocular rivalry. Neuron *42*, 163-172.

Pauling, L., and Coryell, C. (1936). The magnetic properties and structure of hemoglobin, oxyhemglobin and carbon monoxyhemoglobin. Proc Natl Acad Sci U S A *22*, 210-216.

Pedersen, J., Livoir-Petersen, M. F., and Schelde, J. T. (1989). An ethological approach to autism: an analysis of visual behaviour and interpersonal contact in a child versus adult interaction. Acta Psychiatr Scand *80*, 346-355.

Perrett, D. I., Rolls, E. T., and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. Exp Brain Res *47*, 329-342.

Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., and Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. Proc R Soc Lond B Biol Sci *223*, 293-317.

Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., Chitty, A. J., Hietanen, J. K., and Ortega, J. E. (1989). Frameworks of analysis for the neural representation of animate objects and actions. J Exp Biol *146*, 87-113.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., and Thomas, S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. Exp Brain Res *86*, 159-173.

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., Henzi, S. P., Castles, D. L., and Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. Nature *394*, 884-887.

Peskin, M., and Newell, F. N. (2004). Familiarity breeds attraction: effects of exposure on the attractiveness of typical and distinctive faces. Perception *33*, 147-157.

Pessoa, L., McKenna, M., Gutierrez, E., and Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. Proc Natl Acad Sci USA *99*, 11458-11463.

Petersen, S. E., Fox, P. T., Snyder, A. Z., and Raichle, M. E. (1990). Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. Science *249*, 1041-1044.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., and Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. J Cogn Neurosci *12*, 729-738.

Phelps, E. A., Cannistraci, C. J., and Cunningham, W. A. (2003). Intact performance on an indirect measure of race bias following amygdala damage. Neuropsychologia *41*, 203-208.

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., Bullmore, E. T., Perrett, D. I., Rowland, D., Williams, S. C., *et al.* (1997). A specific neural substrate for perceiving facial expressions of disgust. Nature *389*, 495-498.

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., Williams, S. C., Bullmore, E. T., Brammer, M., and Gray, J. A. (1998). Neural responses to facial and vocal expressions of fear and disgust. Proc R Soc Lond B Biol Sci *265*, 1809-1817.

Phillips, M. L., Williams, L., Senior, C., Bullmore, E. T., Brammer, M. J., Andrew, C., Williams, S. C., and David, A. S. (1999). A differential neural response to threatening and non-threatening negative facial expressions in paranoid and non-paranoid schizophrenics. Psychiatry Res *92*, 11-31.

Phillips, M. L., Medford, N., Young, A. W., Williams, L., Williams, S. C., Bullmore, E. T., Gray, J. A., and Brammer, M. J. (2001). Time courses of left and right amygdalar responses to fearful facial expressions. Hum Brain Mapp *12*, 193-202.

Phillips, M. L., Williams, L. M., Heining, M., Herba, C. M., Russell, T., Andrew, C., Bullmore, E. T., Brammer, M. J., Williams, S. C., Morgan, M., *et al.* (2004).

Differential neural responses to overt and covert presentations of facial expressions of fear and disgust. Neuroimage *21*, 1484-1496.

Pierce, K., Muller, R. A., Ambrose, J., Allen, G., and Courchesne, E. (2001). Face processing occurs outside the fusiform 'face area' in autism: evidence from functional MRI. Brain *124*, 2059-2073.

Pixley, E. L., and Bachevalier, J. (2002). Amygdala and orbital frontal cortex lesions differently affect the processing of facial emotions in monkeys. Soc Neurosci Abs *Program No. 86.15.*

Pizzagalli, D. A., Lehmann, D., Hendrick, A. M., Regard, M., Pascual-Marqui, R. D., and Davidson, R. J. (2002). Affective judgments of faces modulate early activity (approximately 160 ms) within the fusiform gyri. Neuroimage *16*, 663-677.

Pizzagalli, D. A., Greischar, L. L., and Davidson, R. J. (2003). Spatio-temporal dynamics of brain mechanisms in aversive classical conditioning: high-density event-related potential and brain electrical tomography analyses. Neuropsychologia *41*, 184-194.

Ploghaus, A., Tracey, I., Gati, J. S., Clare, S., Menon, R. S., Matthews, P. M., and Rawlins, J. N. (1999). Dissociating pain from its anticipation in the human brain. Science *284*, 1979-1981.

Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? Behav Brain Sci *1*, 515-526.

Price, C., Wise, R., Ramsay, S., Friston, K., Howard, D., Patterson, K., and Frackowiak, R. (1992). Regional response differences within the human auditory cortex when listening to words. Neurosci Lett *146*, 179-182.

Price, C. J., and Friston, K. J. (1997). Cognitive conjunction: a new approach to brain activation experiments. Neuroimage *5*, 261-270.

Price, J. L., Carmichael, S. T., and Drevets, W. C. (1996). Networks related to the orbital and medial prefrontal cortex; a substrate for emotional behavior? Prog Brain Res *107*, 523-536.

Puce, A., Allison, T., Gore, J. C., and McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. J Neurophysiol *74*, 1192-1199.

Puce, A., Allison, T., Asgari, M., Gore, J. C., and McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. J Neurosci *16*, 5205-5215.

Puce, A., Allison, T., Bentin, S., Gore, J. C., and McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. J Neurosci *18*, 2188-2199.

Puce, A., Allison, T., and McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. Cereb Cortex *9*, 445-458.

Puce, A., and Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. Philos Trans R Soc Lond B Biol Sci *358*, 435-445.

Purcell, E. M., Torrey, H. C., and Pound, R. V. (1946). Resonance absorption by nuclear magnetic moments in a solid. Physical Review *69*, 37-38.

Quaglino, A. (1867). Emiplegia sinistra con amaurosi – Guarigione – Perdita totale della percezione dei colori e della memoria della configurazione degli oggetti. Giornale d'Oftalmologia Italiano *10*, 106-112.

Quaglino, A., Borelli, G. B., Della Sala, S., and Young, A. W. (2003). Quaglino's 1867 case of prosopagnosia. Cortex *39*, 533-540.

Rafal, R., Smith, J., Krantz, J., Cohen, A., and Brennan, C. (1990). Extrageniculate vision in hemianopic humans: saccade inhibition by signals in the blind field. Science *250*, 118-121.

Raichle, M. E. (1998). Behind the scenes of functional brain imaging: a historical and physiological perspective. Proc Natl Acad Sci U S A *95*, 765-772.

Rapcsak, S. Z., Comer, J. F., and Rubens, A. B. (1993). Anomia for facial expressions: neuropsychological mechanisms and anatomical correlates. Brain Lang *45*, 233-252.

Rapcsak, S. Z., Galper, S. R., Comer, J. F., Reminger, S. L., Nielsen, L., Kaszniak, A. W., Verfaellie, M., Laguna, J. F., Labiner, D. M., and Cohen, R. A. (2000). Fear recognition deficits after focal brain damage: a cautionary note. Neurology *54*, 575-581.

Rees, G., Friston, K., and Koch, C. (2000). A direct quantitative relationship between the functional properties of human and macaque V5. Nat Neurosci *3*, 716-723.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996a). Premotor cortex and the recognition of motor actions. Brain Res Cogn Brain Res *3*, 131-141.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., and Fazio, F. (1996b). Localization of grasp representations in humans by PET: 1. Observation versus execution. Exp Brain Res *111*, 246-252.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. Nat Rev Neurosci *2*, 661-670.

Robinson, S., Windischberger, C., Rauscher, A., and Moser, E. (2004). Optimized 3 T EPI of the amygdalae. Neuroimage *22*, 203-210.

Rolls, E. T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. Hum Neurobiol *3*, 209-222.

Rolls, E. T. (1996). The orbitofrontal cortex. Philos Trans R Soc Lond B Biol Sci *351*, 1433-1443.

Rolls, E. T., Critchley, H. D., Mason, R., and Wakeman, E. A. (1996). Orbitofrontal cortex neurons: role in olfactory and visual association learning. J Neurophysiol *75*, 1970-1981.

Rolls, E. T. (1999). The Brain and Emotion, (Oxford, Oxford University Press).

Rolls, E. T. (2000). The orbitofrontal cortex and reward. Cereb Cortex *10*, 284-294.

Rorden, C., and Brett, M. (2000). Stereotaxic display of brain lesions. Behav Neurol *12*, 191-200.

Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., and Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. Brain *126*, 2381-2395.

Rosvold, H. E., Mirsky, A. F., and Pribram, K. H. (1954). Influence of amygdalectomy on social behavior in monkeys. J Comp Physiol Psychol *47*, 173-178.

Rotshtein, P., Malach, R., Hadar, U., Graif, M., and Hendler, T. (2001). Feeling or features: different sensitivity to emotion in high-order visual cortex and amygdala. Neuron *32*, 747-757.

Roy, C., and Sherrington, C. (1890). On the regulation of the blood supply of the brain. J Physiol *11*, 85-100.

Royet, J. P., Hudry, J., Zald, D. H., Godinot, D., Gregoire, M. C., Lavenne, F., Costes, N., and Holley, A. (2001). Functional neuroanatomy of different olfactory judgments. Neuroimage *13*, 506-519.

Rozin, P., and Fallon, A. E. (1987). A perspective on disgust. Psychol Rev *94*, 23-41.

Rozin, P., Lowery, L., and Ebert, R. (1994). Varieties of disgust faces and the structure of disgust. J Pers Soc Psychol *66*, 870-881.

Sahraie, A., Weiskrantz, L., Trevethan, C. T., Cruce, R., and Murray, A. D. (2002). Psychophysical and pupillometric study of spatial channels of visual processing in blindsight. Exp Brain Res *143*, 249-256.

Samson, D., Apperly, I. A., Chiavarino, C., and Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. Nat Neurosci *7*, 499-500.

Sander, D., Grafman, J., and Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. Rev Neurosci *14*, 303-316.

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". Neuroimage *19*, 1835-1842.

Schmolck, H., and Squire, L. R. (2001). Impaired perception of facial emotions following bilateral damage to the anterior temporal lobe. Neuropsychology *15*, 30-38.

Schultz, R. T., Gauthier, I., Klin, A., Fulbright, R. K., Anderson, A. W., Volkmar, F., Skudlarski, P., Lacadie, C., Cohen, D. J., and Gore, J. C. (2000a). Abnormal ventral temporal cortical activity during face discrimination among individuals with autism and Asperger syndrome. Arch Gen Psychiatry *57*, 331-340.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. J Neurophysiol *80*, 1-27.

Schultz, W., Tremblay, L., and Hollerman, J. R. (2000b). Reward processing in primate orbitofrontal cortex and basal ganglia. Cereb Cortex *10*, 272-284.

Schumann, C. M., Hamstra, J., Goodlin-Jones, B. L., Lotspeich, L. J., Kwon, H., Buonocore, M. H., Lammers, C. R., Reiss, A. L., and Amaral, D. G. (2004). The amygdala is enlarged in children but not adolescents with autism; the hippocampus is enlarged at all ages. J Neurosci *24*, 6392-6401.

Schweinberger, S. R., and Soukup, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. J Exp Psychol Hum Percept Perform *24*, 1748-1765.

Schweinberger, S. R., Burton, A. M., and Kelly, S. W. (1999). Asymmetric dependencies in perceiving identity and emotion: experiments with morphed faces. Percept Psychophys *61*, 1102-1115.

Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., and Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. Brain Res Cogn Brain Res *14*, 398-409.

Schweinberger, S. R., and Burton, A. M. (2003). Covert recognition and the neural system for face processing. Cortex *39*, 9-30.

Schyns, P. G., and Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. Psychol Sci *5*, 195-200.

Schyns, P. G., and Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. Cognition *69*, 243-265.

Senju, A., Yaguchi, K., Tojo, Y., and Hasegawa, T. (2003). Eye contact does not facilitate detection in children with autism. Cognition *89*, B43-51.

Senju, A., Tojo, Y., Dairoku, H., and Hasegawa, T. (2004). Reflexive orienting in response to eye gaze and an arrow in children with and without autism. J Child Psychol Psychiatry *45*, 445-458.

Sergent, J., Ohta, S., and MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. Brain *115*, 15-36.

Sergent, J., Ohta, S., MacDonald, B., and Zuck, E. (1994). Segregated Processing of Facial Identity and Emotion in the Human Brain: A PET Study. Visual Cognition *1* 349-369.

Shah, N. J., Marshall, J. C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H. J., and Fink, G. R. (2001). The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. Brain *124*, 804-815.

Shaw, P., Lawrence, E. J., Radbourne, C., Bramham, J., Polkey, C. E., and David, A. S. (2004). The impact of early and late damage to the human amygdala on 'theory of mind' reasoning. Brain *127*, 1535-1548

Shibata, T., Nishijo, H., Tamura, R., Miyamoto, K., Eifuku, S., Endo, S., and Ono, T. (2002). Generators of visual evoked potentials for faces and eyes in the human brain as determined by dipole localization. Brain Topogr *15*, 51-63.

Siegal, M., and Varley, R. (2002). Neural systems involved in "theory of mind". Nat Rev Neurosci *3*, 463-471.

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., and Frith, C. D. (2004a). Brain responses to the acquired moral status of faces. Neuron *41*, 653-662.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004b). Empathy for pain involves the affective but not sensory components of pain. Science *303*, 1157-1162.

Skottun, B. C. (2000). The magnocellular deficit theory of dyslexia: the evidence from contrast sensitivity. Vision Res *40*, 111-127.

Small, D. M., Gregory, M. D., Mak, Y. E., Gitelman, D., Mesulam, M. M., and Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. Neuron *39*, 701-711.

Sprengelmeyer, R., Young, A. W., Schroeder, U., Grossenbacher, P. G., Federlein, J., Buttner, T., and Przuntek, H. (1999). Knowing no fear. Proc R Soc Lond B Biol Sci *266*, 2451-2456.

Stone, V. E., Cosmides, L., Tooby, J., Kroll, N., and Knight, R. T. (2002). Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. Proc Natl Acad Sci U S A *99*, 11531-11536.

Stone, V. E., Baron-Cohen, S., Calder, A., Keane, J., and Young, A. (2003). Acquired theory of mind impairments in individuals with bilateral amygdala lesions. Neuropsychologia *41*, 209-220.

Strange, B. A., Henson, R. N., Friston, K. J., and Dolan, R. J. (2000). Brain mechanisms for detecting perceptual, semantic, and emotional deviance. Neuroimage *12*, 425-433.

Streit, M., Ioannides, A. A., Liu, L., Wolwer, W., Dammers, J., Gross, J., Gaebel, W., and Muller-Gartner, H. W. (1999). Neurophysiological correlates of the recognition of facial expressions of emotion as revealed by magnetoencephalography. Brain Res Cogn Brain Res *7*, 481-491.

Sugase, Y., Yamane, S., Ueno, S., and Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. Nature *400*, 869-873.

Surguladze, S. A., Brammer, M. J., Young, A. W., Andrew, C., Travis, M. J., Williams, S. C., and Phillips, M. L. (2003). A preferential increase in the extrastriate response to signals of danger. Neuroimage *19*, 1317-1328.

Swanson, L. W., and Petrovich, G. D. (1998). What is the amygdala? Trends Neurosci *21*, 323-331.

Talairach, J., and Tournoux, P. (1988). Co-planar Stereotaxic Atlas of the Human Brain (Stuttgart, Germany, Theime).

Tarr, M. J., and Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. Nat Neurosci *3*, 764-769.

Teves, D., Videen, T. O., Cryer, P. E., and Powers, W. J. (2004). Activation of human medial prefrontal cortex during autonomic responses to hypoglycemia. Proc Natl Acad Sci U S A *101*, 6217-6221.

Thornhill, R., and Gangestad, S. W. (1999). Facial attractiveness. Trends Cogn Sci *3*, 452-460.

Tiberghien, G., Baudouin, J. Y., Guillaume, F., and Montoute, T. (2003). Should the temporal cortex be chopped in two? Cortex *39*, 121-126.

Todorov, A., and Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: evidence from a false recognition paradigm. J Pers Soc Psychol *83*, 1051-1065.

Tolhurst, D. J. (1975). Sustained and transient channels in human vision. Vision Res *15*, 1151-1155.

Tranel, D., Damasio, A. R., and Damasio, H. (1988). Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. Neurology *38*, 690-696.

Tranel, D., Damasio, H., and Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. Neuropsychologia *35*, 1319-1327.

Ugurbil, K., Hu, X., Chen, W., Zhu, X. H., Kim, S. G., and Georgopoulos, A. (1999). Functional mapping in the human brain using high magnetic fields. Philos Trans R Soc Lond B Biol Sci *354*, 1195-1213.

Ungless, M. A., Magill, P. J., and Bolam, J. P. (2004). Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. Science *303*, 2040-2042.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, P., Maier, W., Shah, N. J., Fink, G. R., and Zilles, K. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. Neuroimage *14*, 170-181.

Vogels, R., Biederman, I., Bar, M., and Lorincz, A. (2001). Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. J Cogn Neurosci *13*, 444-453.

Vuilleumier, P., Armony, J. L., Driver, J., and Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. Neuron *30*, 829-841.

Vuilleumier, P., Armony, J. L., Clarke, K., Husain, M., Driver, J., and Dolan, R. J. (2002). Neural response to emotional faces with and without awareness: event-related fMRI in a parietal patient with visual extinction and spatial neglect. Neuropsychologia *40*, 2156-2166.

Vuilleumier, P., Armony, J. L., Driver, J., and Dolan, R. J. (2003a). Distinct spatial frequency sensitivities for processing faces and emotional expressions. Nat Neurosci *6*, 624-631.

Vuilleumier, P., Mohr, C., Valenza, N., Wetzel, C., and Landis, T. (2003b). Hyperfamiliarity for unknown faces after left lateral temporo-occipital venous infarction: a double dissociation with prosopagnosia. Brain *126*, 889-907.

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., and Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. Nat Neurosci *7*, 1271-1278.

Wada, Y., and Yamamoto, T. (2001). Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region. J Neurol Neurosurg Psychiatry *71*, 254-257.

Wason, P., and Johnson-Laird, P. (1972). Psychology of Reasoning: Structure and Content (Cambridge, MA, Harvard University Press).

Weiskrantz, L. (1956). Behavioural changes associated with ablations of the amygdaloid complex in monkeys. J Comp Physiol Psychol *49*, 381-391.

Whalen, P. J. (1998). Fear, vigilance and ambiguity: initial neuroimaging studies of the human amygdala. Curr Dir Psychol Sci *7*, 177-188.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., and Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. J Neurosci *18*, 411-418.

Wicker, B., Michel, F., Henaff, M. A., and Decety, J. (1998). Brain regions involved in the perception of gaze: a PET study. Neuroimage *8*, 221-227.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. Neuron *40*, 655-664.

Wigan, A. L. (1844). The duality of the mind: Proved by the structure, functions, and diseases of the brain, and by the phenomena of mental derangement, and shown to be essential to moral responsibility (London, Longman, Brown, Green, and Longmans).

Williams, J. H., Whiten, A., Suddendorf, T., and Perrett, D. I. (2001). Imitation, mirror neurons and autism. Neurosci Biobehav Rev *25*, 287-295.

Williams, M. A., Morris, A. P., McGlone, F., Abbott, D. F., and Mattingley, J. B. (2004). Amygdala responses to fearful and happy facial expressions under conditions of binocular suppression. J Neurosci *24*, 2898-2904.

Wilson, M., and Daly, M. (2004). Do pretty women inspire men to discount the future? Biol Lett *271*, 177-179.

Wojciulik, E., Kanwisher, N., and Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. J Neurophysiol *79*, 1574-1578.

Wood, J. N. (2003). Social cognition and the prefrontal cortex. Behav Cogn Neurosci Rev *2*, 97-114.

Wood, J. N., Romero, S. G., Makale, M., and Grafman, J. (2003). Category-specific representations of social and nonsocial knowledge in the human prefrontal cortex. J Cogn Neurosci *15*, 236-248.

Worsley, K., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., and Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp *4*, 58-73.

Worsley, K. J., Evans, A. C., Marrett, S., and Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. J Cereb Blood Flow Metab *12*, 900-918.

Worsley, K. J., and Friston, K. J. (1995). Analysis of fMRI time-series revisited--again. Neuroimage *2*, 173-181.

Worsley, K. J., and Friston, K. J. (2000). A test for a conjunction. Statistics and Probability Letters *47*, 135-140.

Wright, C. I., Fischer, H., Whalen, P. J., McInerney, S. C., Shin, L. M., and Rauch, S. L. (2001). Differential prefrontal cortex and amygdala habituation to repeatedly presented emotional stimuli. Neuroreport *12*, 379-383.

Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., and Kanazawa, S. (2003). You can judge a book by its cover: Evidence that cheaters may look different from cooperators. Evol Hum Behav *24*, 290-301.

Young, A. W., Newcombe, F., de Haan, E. H., Small, M., and Hay, D. C. (1993). Face perception after brain injury. Selective impairments affecting identity and expression. Brain *116*, 941-959.

Young, A. W., Aggleton, J. P., Hellawell, D. J., Johnson, M., Broks, P., and Hanley, J. R. (1995). Face processing impairments after amygdalotomy. Brain *118*, 15-24.

Young, A. W., Hellawell, D. J., Van De Wal, C., and Johnson, M. (1996). Facial expression processing after amygdalotomy. Neuropsychologia *34*, 31-39.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., and Perrett, D. I. (1997). Facial expression megamix: tests of dimensional and category accounts of emotion recognition. Cognition *63*, 271-313.

Zald, D. H., Mattson, D. L., and Pardo, J. V. (2002). Brain activity in ventromedial prefrontal cortex correlates with individual differences in negative affect. Proc Natl Acad Sci U S A *99*, 2450-2454.

Zald, D. H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. Brain Res Brain Res Rev *41*, 88-123.

Zarahn, E., Aguirre, G. K., and D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. Neuroimage *5*, 179-197.

Zatorre, R. J., Jones-Gotman, M., and Rouby, C. (2000). Neural mechanisms involved in odor pleasantness and intensity judgments. Neuroreport *11*, 2711-2716.

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., and Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. J Neurosci *11*, 641-649.

Zysset, S., Huber, O., Ferstl, E., and von Cramon, D. Y. (2002). The anterior frontomedian cortex and evaluative judgment: an fMRI study. Neuroimage *15*, 983-991.